ULISES Y LAS SIRENAS

ESTUDIOS SOBRE RACIONALIDAD E IRRACIONALIDAD

JON ELSTER

BREVIARIOS
Fondo de Cultura Económica

Traducción de

JUAN JOSÉ UTRILLA

ULISES Y LAS SIRENAS

Estudios sobre racionalidad e irracionalidad

por

JON ELSTER

FONDO DE CULTURA ECONÓMICA
MÉXICO

Primera edición en inglés, 1979
Segunda edición en inglés, 1984
Primera edición en español, 1989
Segunda reimpresión, 1997

Se prohíbe la reproducción total o parcial de esta obra
—incluido el diseño tipográfico y de portada—,
sea cual fuere el medio, electrónico o mecánico,
sin el consentimiento por escrito del editor.

Título original:
Ulysses and the Sirens. Studies in Rationality and Irrationality
© 1984, Cambridge University Press, Cambridge
ISBN 0-521-26984-9

D. R. © 1989, Fondo de Cultura Económica, S. A. de C. V.
D. R. © 1995, Fondo de Cultura Económica
Carretera Picacho-Ajusco 227; 14200 México, D. F.

ISBN 968-16-4659-2 (Rústica)
ISBN 968-16-3376-8 (Empastado)

Impreso en México

PREFACIO Y AGRADECIMIENTOS

Muchos filósofos y científicos sociales, en algún momento de su vida, desearon escribir literatura o poesía, sólo para descubrir que carecían de lo necesario para realizar esa labor. Otros han escogido la filosofía o la ciencia social como segunda opción, al ver que su primer objetivo, las matemáticas, no estaba realmente al alcance de sus habilidades. La presente obra halla lugar en la intersección de estas dos incapacidades; pero fracasar siempre es fracasar *en algo*, lo que nos deja con un conocimiento de la clase de cosa que, sin éxito, hemos intentado hacer. En los ensayos aquí reunidos he tratado de explotar este conocimiento para hacer un análisis del comportamiento racional y del irracional.

Los ensayos se escribieron de modo independiente unos de otros, sin importar lo cual los he vuelto a redactar para evitar redundancias e incorporar nuevas reflexiones. Tal vez sea útil explicar aquí cómo están relacionados estos temas. El capítulo I establece el paradigma de la conducta individualmente racional, que se distingue a la vez de la adaptación biológica y de la adaptación funcional en las sociedades. La principal idea que aquí defiendo es la siguiente: la racionalidad específicamente *humana* se caracteriza por la capacidad de relacionarse con el futuro, en comparación con el miope y gradual ascenso de la selección natural. El capítulo II, entonces, introduce la idea de conducta imperfectamente racional, cuya necesidad surge porque la flaqueza de voluntad puede impedirnos *utilizar* nuestra capacidad de una con-

ducta perfectamente racional. La idea de *atarse a sí mismo,* como lo hizo Ulises antes de enfrentarse a las Sirenas, es el concepto fundamental del capítulo, aunque también se analiza la estrategia, distinta, de "apuestas indirectas privadas". El capítulo III es, en esencia, una lista de problemas de la teoría del actor racional, expuesta con la intención de evaluar la capacidad de esta teoría en comparación con los enfoques orientados a las normas o estructuralistas. Llego, por esa senda, a la conclusión de que la teoría del actor racional tiene prioridad lógica sobre sus competidoras, aunque no necesariamente sea mejor para cada caso en particular. En el capítulo IV se seleccionan algunos de los problemas ya analizados en el capítulo III, para una discusión más profunda. En particular, trato de explicar cómo es posible interpretar creencias contradictorias y deseos contradictorios como *significativos* aun si son *irracionales.* Dicho en otras palabras, hay una secuencia descendente de racionalidad perfecta, racionalidad imperfecta, racionalidad problemática e irracionalidad que, pese al muy diverso material incluido, presta a los ensayos una unidad conceptual que justifica el haber sido reunidos en un libro.

Subyace en todos estos detalles una visión particular de la filosofía de la ciencia, que espero poder exponer más plenamente en otra ocasión. Un breve esbozo puede preparar al lector para la comprensión de algunas de las ideas que analizo más adelante.

i) Existen básicamente tres modos de explicación en la ciencia: la causal, la funcional y la intencional.

ii) Todas las ciencias utilizan la explicación causal.

iii) Las ciencias físicas sólo emplean la explicación causal; los principios de menor tiempo y otras formulaciones de variación no son más que útiles analíticos, sin ningún poder explicativo.

iv) En la biología no hay lugar para la explicación intencional. Esta afirmación la defiendo en el capítulo I.

v) En las ciencias sociales no hay lugar para la explicación funcional. Defiendo (y condiciono) esta afirmación en el capítulo I, sec. 5 y en el capítulo II, sec. 8.

vi) En biología puede hacerse una distinción entre la causalidad subfuncional (mutaciones, envejecimiento) y causalidad suprafuncional (efectos de derrame, benéficos o nocivos, de las adaptaciones individuales). Toco brevemente esta distinción en el capítulo I.

vii) En las ciencias sociales puede establecerse una distinción similar entre la causalidad subintencional, y la causalidad suprainteneional. La primera se refiere a los procesos causales que ocurren dentro del individuo, formando o pervirtiendo sus intenciones. Éste es el tema de gran parte de los capítulos II y III. El último se refiere a la interacción causal entre individuos. En mi obra *Logical Society*,[1] que en cierto sentido es un volumen gemelo del presente, analizo con cierta extensión el tema.

viii) Se debe estudiar la conducta animal y humana con las nociones de función y de intención como *ideas reguladoras*. No toda conducta animal es funcional, y no toda conducta

[1] Elster (1978a).

humana es racional o intencional, pero sí existe una *suposición* bien fundada de que, típicamente, esto es lo que ocurre.

El capítulo I fue presentado originalmente en el Cuarto Congreso Internacional de la Organización Internacional para el Estudio del Desarrollo Humano, en París, en 1977. La presente versión se ha beneficiado de los comentarios de Roger Masters, Arthur Stinchcombe y George Williams. Una versión mucho más breve del capítulo II fue presentada inicialmente en el Taller ECPR sobre Teoría Política, en Lovaina, en 1976. Luego, Finn Tschudi me ayudó, señalándome la obra muy similar de George Ainslie, de quien después recibí comentarios, y que me dio acceso a manuscritos inéditos que resultaron muy importantes para el desarrollo de mis ideas. También deseo dar las gracias a Francis Sejersted, Sissel Reichelt, Dagfinn Follesdal, John Perry, Michael Bratman, Amélie Rorty, Peter Hammond, Arthur Stinchcombe y Robert Goodin, por sus críticas y sus consejos. El capítulo III fue presentado en el Séminaire International sur l'Économie Sociologique, París, 1977. Deseo agradecer a John Harsanyi sus estimulantes discusiones, durante el periodo de gestación del escrito, y a Robert Goodin sus constructivas críticas. El capítulo IV se traslapa, en parte, con el capítulo IV de mi obra *Logic and Society*, pero hay grandes diferencias, en el material mismo y en la forma en que está organizado. Entre las personas con quienes estoy en deuda se cuentan Amélie Rorty, Eugène Genovese y Paul Watzlawick, la primera por sus importantes sugestiones y los dos últimos por confirmarme que, en realidad, los había yo comprendido correctamente.

I. LA RACIONALIDAD PERFECTA: MÁS ALLÁ DEL ASCENSO DE LA GRADIENTE

1. Introducción

La intención de este capítulo es en gran parte polémica y crítica. Aduciré que pese a ciertas analogías superficiales que existen entre las ciencias sociales y biológicas, hay diferencias fundamentales que hacen muy improbable que tengan mucho que aprender unas de las otras. La diferencia radica, en esencia, en la distinción entre las explicaciones *intencionales* utilizadas en las ciencias sociales y las explicaciones *funcionalistas* que son propias de la biología. Donald Davidson [1] ha argüido que los intentos de reducción psicofísica están condenados a tropezar con el obstáculo de la intencionalidad, y desde un punto de vista muy diferente presentamos aquí una tesis similar con respecto a la reducción o analogía psicobiológica. La discusión se desarrolla, a la vez, contra el uso

[1] Davidson (1973). El rasgo característico de la opinión de Davidson es su simultaneidad materialista, determinista y antirreduccionista. El siguiente pasaje muestra la esencia de su opinión: "Si cierto concepto psicológico se aplica a un hecho y no a otro, entonces debe haber una diferencia descriptible en términos físicos. Pero de ahí no se sigue que hay una sola diferencia físicamente distinguible que reconoce cualesquiera dos hechos que difieren en un respecto psicológico dado" (p. 717). Es decir, concluir, partiendo del determinismo y del materialismo, en el reduccionismo es cometer la misma falacia que cuando del hecho de que todo tiene una causa concluimos que hay algo que es la causa de todo.

de métodos sociológicos en la biología y contra la transferencia de paradigmas biológicos a las ciencias sociales. La mayor parte del espacio está dedicado al primer problema, pues es el menos extensamente analizado en la bibliografía especializada. Yo afirmo que aun si la selección natural puede estimular hasta cierto punto la intencionalidad, existen diferencias críticas en la esencia misma de la adaptación animal y la humana. Dos de tales diferencias son analizadas con cierto detalle. En las secciones 2 y 3 del presente capítulo se plantea un argumento en favor de que la capacidad de la *maximización global* es un rasgo específicamente humano que no se encuentra en la selección natural. En la sección 4 se establece un argumento similar, con respecto a la capacidad exclusivamente humana de *conducta estratégica.* En la sección 5 enfoco el segundo problema, llamando la atención hacia las dificultades mejor conocidas del análisis funcionalista en la sociología. En pocas palabras, mi argumento es, aquí, que en las sociedades no hay un mecanismo general —correspondiente a la selección natural— que nos permita inferir que las funciones latentes de una estructura pueden mantener en forma característica la estructura por obra de la retroalimentación.

El estudio de estas cuestiones me ha mostrado que mi propio argumento se presta a tres equívocos. En primer lugar, no niego que la investigación interdisciplinaria, que reúna los recursos empíricos y teóricos de la biología y de la sociología pueda ser realmente fértil. Por ejemplo, en la evolución humana, la estructura social es parte del medio que determina que una mutación dada sea benéfica o nociva. A otro nivel, los componentes biológicos del lenguaje, inteligencia o en-

fermedad mental están hoy bien documentados, aun si la importancia relativa de naturaleza y cultura sigue siendo materia de controversia. Mi argumento sólo se dirige a la transferencia de paradigmas explicativos enteros. Al recurso de las intenciones en biología y de las funciones en sociología.

En segundo lugar, no niego que puedan encontrarse casos en que la adaptación biológica conduce a una desviación de los máximos locales, y al alcance de los máximos globales, ni que la explicación funcional a veces pueda ser útil en sociología. Me limito a argüir que no puede haber una suposición *general* de que los máximos globales serán alcanzados por selección natural, y que las funciones latentes no explican de *forma característica* la persistencia de las estructuras sociales que muestran estas funciones. Podemos recordar aquí el concepto kantiano de *idea reguladora*, en oposición a idea dogmática.[2] Una idea reguladora distribuye la carga de la prueba, en el sentido de postular los rasgos que pueden presumirse como hipótesis de trabajo, a falta de prueba específica en sentido contrario, y qué rasgos deberán mostrarse como presentes en cada caso en particular. La afirmación "todo tiene una función" no debe

[2] *Kritik der reinen Vernunft*, B 670 *ss* (A642 *ss*); *Kritik der Urteilskraft*, § 70 *ss*. El ejemplo más sencillo tal vez sea la reinterpretación kantiana del "principio de continuidad" partiendo de una aserción dogmática en el sentido de que el conjunto de formas orgánicas puede ser ordenado en una serie continua (¿o densa?), a un principio regulador que nos exhorta a buscar, entre cualquiera de las formas, una intermedia. La afirmación "todo tiene una causa" es en la filosofía kantiana un principio constitutivo del que no puede haber excepción, mientras que la idea de que "todo tiene una función" es un principio regulador que sí admite excepciones, como queda explicado en la n. 83, *infra*.

tomarse como planteamiento dogmático acerca de los organismos,[3] sino como afirmación en el sentido de que podemos presuponer legítimamente que un rasgo dado tiene una función —es decir, no podría sufrir pequeñas variaciones sin la pérdida de la capacidad reproductiva para el organismo en cuestión— hasta que se demuestre lo contrario. Desde luego, la justificación de esta idea reguladora es el mecanismo general de ascender gradientes por selección natural, como se explica en la sección 2. Por otra parte, no hay nada que pueda justificar la idea reguladora de que cada rasgo del organismo alcanza un máximo *global*, de modo que todas las variaciones —pequeñas o grandes— implicaran una pérdida de la capacidad reproductiva. Para la adaptación intencional, lo cierto es precisamente lo contrario. Aquí contamos con un mecanismo general para alcanzar máximos globales, y lo que necesita una explicación separada es, en realidad, el no lograrlo.

En tercer lugar, no quiero ser interpretado como antirreduccionista, al menos, no como miembro de la extrema variedad que postula una brecha insalvable entre la adaptación animal y la humana. Creo que la capacidad humana de maximización global debe ser explicada, a la postre, como resultado de la maximización local de ascender gradientes, en la selección natural. En este

[3] Para una declaración comprensible de esta visión dogmática (o que no es no darviniana) podemos tomar la siguiente: "Una 'situación sin función' es sencillamente una cuya función aún no se ha determinado" (Mayr 1970, p. 127). Una declaración según los lineamientos esbozados en el presente ensayo es la siguiente: "El estudio experimental de adaptación ha desenredado valores adaptativos en detalles tan discretos e inconspicuos de organización organísmica que pensaríamos en una característica que tuviese valor de supervivencia hasta que se demostrara lo contrario" (Curio 1973, p. 1049).

sentido, en verdad soy reduccionista; obsérvese, sin embargo, que el objeto de la reducción es la *capacidad* de comportarse de esta manera, y no unos casos específicos en que se ejerza la capacidad. Es decir, creo que la conducta racional debe reducirse en dos pasos: primero, subsumiéndola bajo la capacidad general de resolver problemas racionalmente y, en segundo lugar, explicando esa capacidad *general* por el funcionamiento de la selección natural. El argumento es similar al que propusieron Peter Richerson y Robert Boyd [4] con respecto a los fundamentos biológicos de la *capacidad* de cultura, que en su modelo es muy diferente de los fundamentos biológicos para la *conducta* cultural. En contraste, el argumento es claramente incompatible con las ideas propuestas por C. D. Darlington [5] o R. D. Masters,[6] que suelen ver las pautas conductuales *específicas* en perspectivas biológicas. En la sección 3 volveré a la distinción general-específico.

Como última advertencia, deseo mostrar lo que, según espero, puede resultar una profecía que vaya contra sí misma. No creo que mis críticas, si alguien las lee, vayan a ejercer gran repercusión. En particular, la atracción de las analogías biológicas para los científicos sociales parece ser tan grande que hasta los mejores espíritus se dejan despistar. En la galería de horrores del pensamiento científico hay espacio, ¡ay! no sólo para un Worms o un Lilienfeld, sino también para un Durkheim o un Merton.[7] En el momento actual, creo que podemos predecir, con toda confianza,

[4] Richerson y Boyd.
[5] Darlington (1969).
[6] Masters (1975, 1976).
[7] Para ejemplos de modelos biológicos en las ciencias sociales, *véanse* Stark (1962), Schalenger (1971), Banton (comp.) (1961).

que la etología y la sociobiología generarán, entre ellas, una corriente continua de escritos pseudocientíficos durante varios años. Desde luego, no estoy negando la utilidad ocasional de las analogías biológicas como fuente de nuevas hipótesis, así como no niego que algunos científicos puedan sacar sus mejores ideas de una lectura de la Biblia o de la *Dialectica de la Naturaleza*. En cambio, niego que las analogías biológicas deban tener una categoría privilegiada. Las ideas deben juzgarse por sus descendientes, no por sus antepasados. Espero que el científico social que se proponga dedicar su tiempo a descubrir la analogía social de, digamos, la conducta de predominio o de fuga en los animales reflexione, antes, sobre todos lo seudodebates del siglo XIX acerca de la analogía social correcta de la célula: ¿el individuo o la familia? No hay ninguna razón para que exista semejante analogía, y a esto me limito exclusivamente.

2. La máquina maximizadora local

En esta sección esbozo una versión en extremo simplificada de la teoría de la selección natural. Ateniéndome estrictamente a los primeros principios, espero no equivocarme de manera demasiado patente en cuestiones específicamente biológicas que estén fuera de mi competencia. Como las propias conclusiones estarán al nivel de los primeros principios, creo que se justifican mis simplificaciones,

Concibo, pues, los organismos de una población como una máquina que recibe inductos [*inputs*] en forma de mutaciones. Para simplificar las cosas, presuponemos reproducción asexual, de

modo que las mutaciones son la única fuente de novedad genética. De otra manera, podemos evitar la recombinación arguyendo que, a largo plazo, sólo las mutaciones pueden perturbar el equilibrio biológico (en un medio constante). La corriente de inductos es aleatoria, en el sentido de que no hay correlación entre los requerimientos funcionales del organismo y la probabilidad de ocurrencia de una mutación que satisfaga estos requerimientos. Mediante mutágenos, como los rayos X es posible aumentar la probabilidad de mutaciones en general; por medio de mutágenos químicos se puede aumentar la probabilidad de subgrupos estructuralmente definidos de mutaciones, como la probabilidad de que la citosina se convierta en quinina como resultado de una deaminación por ácido nitroso; pero nunca es posible —y éste es el *dogma central* de la biología molecular— modificar la probabilidad de subgrupos funcionalmente definidos de mutaciones. Aquí puede ser útil una analogía. Comparando las mutaciones con errores de imprenta, podemos aumentar la probabilidad de los errores de imprenta rompiendo los espejuelos del tipógrafo, pero no hay manera de aumentar selectivamente la probabilidad de que ocurran erratas en la segunda edición de un libro que corrija los errores reales de la primera edición.

Presuponemos —decisivamente— que todas las mutaciones son pequeñas, siendo sustituciones características de aminoácidos resultantes de la errata de una sola letra en el código genético. No hay mecanismos de duda, como duplicación de genes, que puedan producir macromutaciones, pero en primer lugar la importancia evolutiva de éstas se halla muy lejos de ser clara y, en segundo lugar, tales mutaciones, aunque grandes si se comparan

con sustituciones de aminoácidos, son pequeñas en comparación con las discontinuidades que se encuentran en la adaptación humana.[8] Ninguna duplicación de genes podría producir un cambio del orden de magnitud del cambio del carro tirado por caballos al carro "sin caballos".[9] Dado que la primera afirmación de este capítulo es que existe una diferencia básica entre la optimización local por medio de pequeñas mejoras y la maximización global que permite dar pasos de cualquier magnitud, en realidad no tiene importancia la definición precisa de "pequeño".

Según la frase de Jacques Monod, la selección natural opera por azar y necesidad.[10] Aunque las mutaciones sean al azar, el proceso de selección es determinista, en el sentido de que la máquina en cualquier momento tiene normas bien definidas para aceptar o rechazar cualquier mutación

[8] Para el primer punto, *véase* Frazzetta (1975), pp. 93 *ss*; para el segundo, *ibid.*, pp. 20, 152. Para una posible consecuencia evolutiva (de *no función*) de duplicación de genes, *véase* Rigby *et al.* (1974), quienes arguyen que el gene duplicado puede mutar a uno funcionalmente inactivo (un paso atrás) y luego mutar a una forma viable y posiblemente superior (dos pasos adelante).

[9] Frazzeta (1975), p. 152. *Cf.* también Schumpeter (1934), p. 54: "Añádanse sucesivamente todos los carruajes del correo que se quieran; no se logrará con ello hacer un ferrocarril." Para otra opinión del desarrollo tecnológico, que subraya los pequeños cambios acumulativos, *véanse* Rosenberg (1976), pp. 66, 166 y *passim*; David (1975); Nelson y Winter (1974, 1976); *cf.* también n. 27 y 28 *infra*. Sin embargo, aun si aceptamos que muchos cambios evolutivos son mayores que las mutaciones de un solo paso, y que muchos casos de cambio técnico son acumulativos, en lugar de discontinuos, creo yo que queda una diferencia genuina que basta para mi argumento. Otra variedad importante (Frazzetta 1975, p. 20) es que en la evolución no hay nada correspondiente a los "fracasos útiles" en la ingeniería. La evolución nunca aprende de pasados errores.

[10] Monod (1970); *véase* también Schoffeniels (1973).

dada. (Esto significa que no trataré con ningún propósito genético ni con el problema de la evolución no darwiniana). La mutación es aceptada si el primer organismo en que ocurre se beneficia en la forma de mayor capacidad reproductiva. Dado que el organismo dejará más descendientes que otros organismos, el nuevo alelo se difundirá por la población hasta encontrarse universalmente presente. (Esta afirmación implica que no trataré de la selección dependiente de frecuencia ni de otras fuentes —como la heterosis— del polimorfismo estable. Sin embargo, véase la sección 4 para un análisis del polimorfismo en su relación con estrategias mixtas.) Entre los resultados de una mutación aceptada vemos que cambiarán las normas para aceptar o rechazar nuevas mutaciones. El organismo se encuentra ahora en un estado distinto del que se encontraba antes de la mutación, y puede ser beneficiado o dañado por diferentes insumos. En pocas palabras: la máquina dice Sí o No a cada inducto de acuerdo con normas que cambian cada vez que dice Sí. Si la máquina llega alguna vez a un estado en que diga No a cada uno de los posibles inductos (de número finito), decimos que ha llegado a un máximo local. La población asciende a lo largo de una gradiente de aptitud hasta llegar a un punto del cual todo nuevo movimiento sólo puede ser descendente; y allí llega a un alto. Para un estado inicial determinado pueden ser accesibles varios máximos locales, y la elección entre ellos depende del orden aleatorio en que ocurren las mutaciones.

Otro nuevo análisis de esta *máquina de maximización local* deberá tomar en cuenta la posibilidad del cambio ambiental. Si el medio cambia, también cambiarán las normas para decir Sí o No a las mutaciones. Una mutación no es benéfica

ni dañina en sí misma sólo con respecto a un trasfondo genético dado (que a su vez es resultado de mutaciones previas) y un medio dado. Con un medio cambiante bien puede ocurrir que ni siquiera sean alcanzados jamás los máximos locales instantáneos, si el organismo no puede mantener el ritmo del medio circundante. La idea de "medio" es, empero, ambigua. En primer lugar, los cambios ambientales pueden referirse a cambios de las condiciones geológicas o climáticas en la medida en que causan un cambio evolutivo sin ser, a su vez, afectadas por él. (Esta última condición se necesita para excluir los cambios climáticos endógenos, como los cambios en la atmósfera generados por la evolución de las plantas.) En segundo lugar, algunas partes del medio son, a su vez, organismos en evolución, o bien los efectos de tal evolución. Si una población se ve sometida a un cambio ambiental exógeno en forma constante, nunca podrá alcanzarse un estado estable, pero en la medida en que el propio medio está integrado (o es efecto) de organismos en evolución, tiene sentido preguntar si puede alcanzarse un *equilibrio general*, en que todos los organismos hayan alcanzado máximos locales en su relación mutua.[11]

Las condiciones para la existencia y consecución de un equilibrio biológico general puede suponerse que son muy complejas. Aquí me limitaré a subrayar la idea general de que la evolución puede ser considerada bajo las suposiciones cada vez más complejas de un medio paramétrico, un medio estratégico (o, en vista de la sección 4, casi estratégico), y de un medio exógenamente cambiante. En el modelo abstracto utilizado aquí sólo me enfrento al primer caso, aun cuando creo que

[11] Winter (1971), p. 258.

las conclusiones pueden extenderse con facilidad al segundo caso. Sin embargo, con el tercer caso cobran pertinencia algunas consideraciones por completo distintas. En particular, el fenómeno de la *adaptación previa* —que surge del hecho de que "cualquier estructura tiene propiedades más allá de aquellas para las que fue construida" [12]— permite alcanzar óptimos globales que en un medio constante habrían permanecido inaccesibles. Sin embargo, éste es exclusivamente un fenómeno al azar, y no un mecanismo general comparable al de la selección natural.

Creemos haber dicho lo suficiente para explicar por qué la obtención de máximos locales en el mundo orgánico no es cuestión que pueda darse por sentada. No hay objeción lógica a la idea de un mundo en que el ritmo de cambio del medio por relación a la tasa de mutación es tan alta que, casi todo el tiempo, la mayoría de los organismos están mal adaptados entre sí y hacia su medio inorgánico. Sin embargo, en el mundo que conocemos, las adaptaciones infinitamente sutiles que encontramos en la estructura y la conducta de los organismos son hechos que durante milenios han causado el desconcierto y (con menos justificación) la admiración de los naturalistas. En muchos casos bien documentados, la solución natural a los problemas estructurales y funcionales está notablemente cerca de la selección que habría escogido un ingeniero o un economista que se enfrentara a los mismos problemas. En algunos casos, animales y hombres *sí* se enfrentan a los mismos problemas, por lo que pueden compararse sus soluciones reales. Como lo ha demostrado d'Arcy Wentworth Thompson en su obra ya clásica *On Growth and Form* así como varios

[12] Salthe (1972), p. 9.

autores recientes,[13] estas soluciones a menudo son fuertemente convergentes. En las recientes obras ecológicas,[14] la naturaleza es considerada más como economista que como ingeniero. El presupuesto óptimo, la programación lineal, la maximización de ganancias y la minimización de costos forman hoy parte de la teoría evolutiva, tanto como de la economía.

A veces se dice que la teoría de la selección natural es una tautología; que la supervivencia del más apto significa la supervivencia de los supervivientes. Si por aptitud queremos decir aptitud genética, esto es verdad; en cambio, si interpretamos aptitud como aptitud ecológica, como se le puede medir, por ejemplo, por la duración de vida del organismo, la supervivencia del más apto se convierte en una proposición empírica de la que podemos ofrecer muchos importantes contraejemplos. Para tomar un solo ejemplo, la selección natural funciona en tal forma que produzca el sacrificio óptimo de los padres por sus crías. Un exceso de sacrificio por una cría dada puede reducir la oportunidad de tener después más crías, mientras que demasiado poco sacrificio dañaría la "inversión" ya hecha en la cría. En cambio, si el objetivo era maximizar la adaptación ecológica, el nulo sacrificio sería claramente la estrategia óptima; en realidad, la mejor estrategia sería no tener ninguna reproducción en absoluto. La confusión entre aptitud ecológica y aptitud genética ocurre, en varias formas, en la bibliografía. Un ejemplo bastante craso se encuentra en una obra reciente sobre la filosofía de la biología, cuyo

[13] Leigh (1971), parte I; Frazzetta (1975), cap. v y *passim*. Sobre d'Arcy Thompson, *véase* Gould (1976).

[14] Para críticas, *véanse* Cody (1974) y Rapport y Turner (1977).

autor afirma que "el proceso de reproducción en los seres humanos" muestra que "lo que puede ser útil para el grupo (por ejemplo, una especie) puede no tener ninguna utilidad para el individuo o puede hasta tener utilidad negativa".[15] Un error más sutil (o potencial para el error) subyace en el frecuente empleo del término "inversión parental" en la teoría evolucionista reciente.[16] Mientras que la inversión económica significa el sacrificio actual de algo (es decir, de consumo) para tener después más *de lo mismo*, la inversión parental significa el sacrificio de algo (es decir, aptitud ecológica) para tener *más* de *otra cosa* (es decir, aptitud genética).

3. La máquina maximizadora global

Para los fines de nuestro argumento, los rasgos decisivos de la máquina de maximización local se refieren a lo que *no puede* hacer. En particular, la máquina es incapaz de *esperar* y de emplear *estrategias indirectas*. Estas ideas se definen de esta manera: la máquina es capaz de esperar si puede decir No a una mutación favorable, con objeto de poder decir Sí a otra, aún más favorable, que llegara después. Supongamos que de una

[15] Simon (1971), p. 82.

[16] *Cf.* especialmente Trivers (1972), p. 139, quien define la investigación parental como "*cualquier inversión por el padre en una cría individual que aumenta las posibilidades de supervivencia de la cría (y por tanto, su éxito reproductivo) al costo de la capacidad del padre de invertir en otra cría*". La inversión, en el sentido económico, sería antes bien, "cualquier inversión por el padre en una cría individual que aumenta la posibilidad de la cría de tener crías sobrevivientes (y por tanto, éxito reproductivo) al costo de la posibilidad del padre de tener crías supervivientes".

proteína dada, la forma B es superior a la forma A y la forma C es superior a B, y que, dada la estructura del código genético, son posibles mutaciones de un paso, de A a B y de A a C, pero no de B a C. Si la mutación a B ocurriera primero, el organismo sería incapaz de decir No. La máquina es capaz de estrategias indirectas si puede decir Sí a una mutación desfavorable para poder decir después Sí a una muy favorable. Supongamos que A es superior a B y C a A, y que pueden ocurrir mutaciones de un solo paso, de A a B y de B a C, pero no de A a C. Si surge B, el organismo no será capaz de decir Sí. El organismo, en cuanto asciende gradientes, tiene sus miopes ojos fijos al terreno, y es incapaz de tomar en cuenta lo que ocurre tras la siguiente colina. Como ha dicho un autor, "los acontecimientos o las posibilidades futuras sencillamente no pueden tener ningún efecto sobre las combinaciones genéticas de los organismos".[17] Sin duda es verdad que "hasta las mutaciones á codones que son completamente sinónimas entre sí pueden, en muchos casos, alterar las futuras posibilidades mutacionales del genotipo",[18] pero estas posibilidades no pueden ejercer *ahora* una presión selectiva que diese una ventaja a una mutación aparentemente neutral sobre el *statu quo*.

Por contraste, aguardar el empleo de estrategias indirectas son rasgos decisivos de la elección humana. Sugeriré que el hombre en realidad puede ser considerado como una *máquina maximizadora global*, caracterización que se remonta a Leibniz, quien consideró que el hombre era exclusivamente capaz de "reculer pour mieux sauter".[19] Algu-

[17] Salthe (1972), p. 133.
[18] Stebbins y Lewontin (1972), p. 24.
[19] Leibniz (1875-1890), III, pp. 346-578; VII, p. 568. Estas

nos ejemplos de maximización global de la vida económica y la vida política son:

	Aguardar	*Estrategias indirectas*
Política	Antiactivismo	Antirreformismo
Economía	Sistema patente	Inversión

Tal vez quepa hacer algunos comentarios. La inversión tal vez sea el ejemplo más sencillo de maximización global que exige superar un máximo local: un paso hacia atrás, para poder dar dos pasos adelante. Como ya lo observamos, la inversión parental en el reino animal no es, de hecho, un ejemplo de inversión en este sentido. Por ejemplo, si una generación aceptara un número de crías menor al máximo, para tener un mayor nú-

referencias en realidad invocan estrategias indirectas como parte de la racionalidad *divina* en la construcción del mejor de todos los mundos posibles. También abundan proposiciones similares acerca del *hombre*, como la siguiente: "Les appétitions sont comme la tendance de la pierre qui va le plus droit mais non pas toujours le meilleur chemin vers le centre de la terre, *ne pouvant pas prévoir* que'elle rencontrera des roches où elle se brisera, au lieu qu'elle se serait aprochée davantage de son but, si elle avait eu *l'esprit et le moyen de se détourner*... Le bonheur est donc pour ainsi dire un chemin par des plaisirs; et le plaisir n'est qu'un pas et un avancement vers le bonheur, le plus court qui se peut faire *selon les présentes impressions*, mais non pas toujours le meilleur. On peut manquer le vrai chemin, en voulant suivre le plus court, comme la pierre allant droit, peut rencontrer trop tôt des obstacles, qui l'empêchent d'avancer assez vers le centre de la terre. Ce qui fait connaitre, que c'est la raison et la volonté qui nous mènent vers le bonheur, mais que le sentiment et l'appétit ne nous portent que vers le plaisir" (v, pp. 175, 182; las cursivas son nuestras). *Cf.* Elster (1975a), cap. VI, para mayor análisis.

mero de nietos, ésta sería, en realidad, inversión en el sentido más estricto, pero semejante idea jamás se ha sostenido en la bibliografía de la vida biológica. El sistema patente tiene el rasgo paradójico de que "al hacer más lenta la difusión del proceso técnico, asegura que haya más progreso técnico que difundir".[20] Como lo subrayó Joseph Schumpeter, la explotación máxima de las posibilidades actuales puede ser, a menudo, un obstáculo para la creación máxima de posibilidades nuevas.[21] Como lo sugirió Maurice Meisner, esta proposición también puede aplicarse al activismo como obstáculo a la acción a largo plazo.[22] Una clase revolucionaria debe ser capaz de aguardar, de "hacer tiempo", de decir No a oportunidades favorables. Y a la inversa, se puede ver al reformismo político como la negación de aprovechar estrategias indirectas. Con las técnicas de planeación acumulativa o con la "ingeniería social por partes" [23] de Karl Popper, sólo se alcanzarán máximos locales, mientras que una clase revolucionaria es capaz de decir Sí a posibilidades desfavorables. (Incidentalmente, la distinción entre máximos locales y globales también es pasada por alto en la obra de Popper sobre biología, donde erróneamente afirma que "lo que Darwin nos mostró fue que el mecanismo de la selección natural... puede estimular una acción humana racional dirigida a un propósito o meta".)[24] Huelga decir que estas observaciones son puramente formales y no implican ninguna opinión política en particular. Sólo estoy diciendo que la *capacidad*

[20] Robinson (1956), p. 87.
[21] Schumpeter (1954), p. 87.
[22] Meisner (1967), pp. 166 *ss.*
[23] Popper (1957), pp. 64 *ss.*
[24] Popper (1972), p. 267.

humana para una conducta revolucionaria —para hacer tiempo y utilizar estrategias indirectas— es un rasgo que distingue al hombre del comportamiento activista y reformista (es decir, oportunista) de la selección natural. Como lo afirmaré más extensamente en el capítulo II, más adelante, a menudo no se empleará esta capacidad aun cuando fuera deseable hacerlo; a la inversa, hay, con frecuencia, muy buenas razones contra experimentar con situaciones que estén muy lejos de la actual.[25]

Desde luego, otra diferencia consiste en que el hombre puede aumentar la probabilidad de ocurrencia de algunos subgrupos —funcionalmente definidos— de "mutaciones" (es decir, nuevas ideas, invenciones o instituciones). Como dice Michael Ruse, "podemos hablar de muchos nuevos elementos culturales, si no la mayoría que aparecen por virtud del hecho de que se les necesita".[26] No estamos obligados a sentarnos a aguardar a que ocurra la mutación necesaria; activamente podemos buscar el nuevo tipo de invento o de institución que se necesita. Ahora bien, la importancia relativa del azar y el carácter directo en el desarrollo tecnológico es cuestión de cierta controversia, y al menos un reciente autor ha argüido que las invenciones al azar pueden generar un proceso que estimule la investigación conscientemente dirigida.[27] Sin embargo, su argumento pretende

[25] Una de estas razones está relacionada con el problema de las preferencias endógenamente cambiante, analizadas en la sección 6 del capítulo II, *infra*. Otra razón se debe a la aversión al riesgo; una tercera, al "efecto de irreversibilidad" analizado por Henry (1974).

[26] Ruse (1974), p. 432.

[27] David (1975), cap. I. David se propone explicar el vínculo existente entre la escasez de factores y las tendencias a evitar factores, en el cambio técnico. Rechaza las

explicar el cambio tecnológico en el siglo XIX, y no pretende tener validez universal. En realidad, parece obvio que la gran empresa moderna busca en forma muy activa nuevas técnicas con propiedades sumamente específicas, a diferencia del aprendizaje más miope, haciendo lo que acaso caracterizara a la empresa pequeña el siglo pasado. Asimismo, aun si la *dirección* del cambio tecnológico es más o menos azarosa, el *ritmo* al que procede puede depender decisivamente de la necesidad de invenciones, como en otra obra reciente.[28]

explicaciones "intencionales" propuestas por Hicks (1932), Fellner (1961) y Kennedy (1964), y luego afirma que: "El movimiento de los desarrollos tecnológicos generados a lo largo del tiempo dentro de un medio económico bastante estable debe ser considerado, ante todo, como un fenómeno distintamente *histórico*, dado que puede surgir por medio de selecciones miopes que productores pasados hicieron entre las distintas especies de técnicas con que originalmente se pusieron a trabajar. El elemento de 'guía' presente en el proceso a largo plazo está atado en gran medida a las circunstancias que rodean esas selecciones, e incluye unas respuestas innovativas inducidas, sin ver hacia adelante, a las actuales señales del mercado o a advertencias futuras" (pp. 61-62). Según esta opinión, el progreso técnico (sobre todo por medio de aprender haciendo) es localmente neutral, pero termina creando una tendencia global hacia economizar en factores escasos.

[28] *Véanse* especialmente Nelson y Winter (1976), quienes arguyen que el progreso técnico ocurre por satisfacción y no por maximización (*cf.* también la sección 4 del presente capítulo y la sección 5 del capítulo III, *infra*). Cuando las ganancias brutas caen por debajo de cierto nivel crítico, se inicia una actividad innovadora e imitadora. La investigación es local y sin tendencias (es decir, al azar), pero el resultado puede mostrar pautas específicas, dependiendo del contexto de la investigación. (Hay una incongruencia en el modelo, cuando Nelson y Winter incluyen el aprender haciendo entre los mecanismos de investigación, pues como lo subrayó David [1975], caps. I-III, este tipo de cambio técnico ocurre mediante la sencilla opera-

Como observación más general, también debe notarse que aun si se ha reparado en que la investigación al azar ocurre en las elecciones humanas, esto no significa que, a la postre, la investigación sea generada por un proceso aleatorio. Por lo contrario: un cálculo previo acaso haya descubierto que una investigación al azar constituiría el proceso más racional, tal vez porque una investigación más conscientemente dirigida habría incluido costos inaceptables para adquirir y evaluar la información.[29] En el caso especial del cambio técnico, un administrador bien informado podrá preferir dejar a sus científicos libres de hacer lo que quieran, porque esta política atraerá mejores científicos y en lo fundamental, a la postre, resultará más conveniente que una política en que científicos menos calificados pasen más tiempo en busca de objetivos específicos.[30]

Estas últimas observaciones nos llevan a un

ción del proceso de producción y no requiere de ninguna inducción particular. Otra incongruencia es que Nelson y Winter [1974] consideran su teoría, neoschumpeteriana, pese al poderoso hincapié hecho en la investigación *local*, en contraste con las invenciones discontinuas.) En la selección natural, desde luego, ni el ritmo de mutaciones ni la dirección las mutaciones es respuesta a las necesidades funcionales del organismo; hay una corriente más o menos continua de mutaciones aleatoriamente dirigidas, como en el modelo de David. Sin embargo, la analogía no es completa, pues en el modelo de David la corriente de inventos no es un paseo al azar *tout court*, sino un paseo al azar con barreras elásticas, que impele la corriente *real* de inventos (y no sólo la dirección *esperada* de inventos) hacia la neutralidad. En otras palabras, hay un *efecto de histéresis* (Elster 1976a), al que no hay analogía en la selección natural.

[29] *Cf.* sección 4 del presente capítulo y sección 5 del capítulo III, *infra*, para algunas dudas acerca de este argumento.

[30] *Véanse* Cotgrove y Box (1970), para testimonios sobre este punto.

nivel nuevo de problemas, y requieren algunas distinciones que resultarán decisivas en la secuela. En particular, debemos distinguir entre la elección de una acción y la elección de un método para elegir. Cada una de estas dos elecciones puede hacerse deliberadamente, o por prueba y error, lo que nos deja con cuatro combinaciones posibles. En el párrafo anterior mencioné el caso de una elección deliberada de un proceso aleatorio de búsqueda, pero para los fines de nuestro argumento actual resulta más importante comprender la posibilidad inversa: prueba y error, que generan una elección deliberada e intencional.

En este punto, el lector puede sentirse, con justicia, un tanto confuso, pues, exactamente, ¿cuáles son los fenómenos que estamos comparando? Si estamos contrastando la adaptación animal y la humana, la comparación apropiada, puede suponerse, sería entre la conducta animal y la humana, mientras que nosotros hemos estado organizando una confrontación entre la selección natural y la elección humana. En realidad, participan aquí tres fenómenos diferentes: adaptación animal, adaptación humana y la selección natural que forjó a ambas. Para mostrar la relación existente entre estos tres dominios, utilizaré el instructivo ejemplo de los *juegos de persecución*. En la teoría de los juegos diferenciales existe un ejemplo común llamado el "juego del chofer homicida", donde un automóvil, en un estacionamiento infinito y totalmente libre de obstáculos, trata de atropellar a un peatón.[31] El perseguidor tiene mayor velocidad, pero menor movilidad debido al radio mayor de curvatura del automóvil. Es intuitivamente claro y puede demostrarse rigurosamente, que si al principio el perseguidor está muy cerca del que

[31] *Cf.* Isaacs (1965) y Hájek (1975).

huye, la estrategia óptima para el perseguidor consiste en *apartarse* del que huye y luego, cuando esté lo bastante lejos, dar vuelta en redondo para iniciar una persecución directa. Se requiere un paso hacia atrás para dos pasos adelante. Un ejemplo típico de las estrategias indirectas que el hombre es capaz de usar y que en realidad usa en muchas de sus actividades de la vida real.

En el reino animal se encuentran tales persecuciones en dos niveles. En el nivel individual podemos ver al zorro que persigue a la liebre a través de los campos en cualquier día de otoño. En el nivel de la especie, el zorro está cazando a la liebre a través de las generaciones, adaptándose continuamente a las contraadaptaciones de la última. Diré algunas palabras acerca de la persecución de la especie, y luego me extenderé un poco más hablando de la persecución individual. El resultado de la sección 2 fue que en la persecución de la especie, son imposibles las estrategias indirectas y las esperas (en el modelo sencillo), pues un organismo que dio dos pasos atrás no dejaría descendientes para dar después los dos pasos adelante, mientras que un organismo que se negara a dar un paso adelante es inconcebible (en realidad, desde luego, la deriva genética [32] y la preadaptación invalidan esta proposición, pero éstos son, esencialmente, fenómenos al azar). Así pues, la pregunta consiste en saber si la maximización local de la selección natural es capaz de producir un maximizador global; si el proceso

[32] Wright (1970, pp. 12 *ss*) explica cómo la deriva al azar puede llevar a una población sobre una "silla de máximos locales". A la inversa Wilson (1975, p. 24) observa que la deriva genética puede impedir que la especie jamás alcance un máximo *local*, "si la especie se apoya en un borde agudo, que conduzca, hacia arriba, al pico adaptativo".

de ascender gradientes al nivel de la especie es capaz de producir, por evolución, individuos que pueden levantar la vista del suelo y mirar más allá de la colina próxima. En un sentido abstracto, sabemos que la respuesta debe ser positiva, pues el hombre mismo es, a la vez, un producto de la selección natural y un verdadero muestrario de estrategias indirectas. No obstante, tal vez sea útil mirar más detenidamente el problema tal como surge en el marco del depredador y su presa.

M. Edmunds [33] y E. Curio [34] han estudiado con cierto detalle el comportamiento de la presa y de los depredadores, respectivamente. Al parecer, la presa no utiliza estrategias indirectas, pero es muy capaz de aguardar; los depredadores son capaces de las dos cosas. Por ejemplo, aún si la fuga puede reducir el peligro actual, también puede aumentar el peligro posteriormente, al indicar la existencia de la presa al depredador; por esta razón, la presa ha "aprendido" (por medio de la selección) a no dejarse llevar por el pánico.[35] Un ejemplo simétrico de aguardar en los depredadores es el "trote especulativo y poco llamativo [que] acerca más al león a una presa, que cualquier carrera veloz, pero que por lo mismo, lo pondría en estado de alerta".[36] Asimismo, "toda una variedad de animales son capaces de desviarse mientras acechan" [37] y de interceptar el camino de fuga de su presa.[38] La presa, a su vez, ha creado una contraestrategia para la intercepción. No se trata de la contraestrategia directa de un continuo cambio de dirección para apartarse siempre de su

[33] Edmunds (1974).
[34] Curio (1976).
[35] Edmunds (1974), p. 251.
[36] Curio (1976), p. 133.
[37] *Ibid.*, p. 137.
[38] *Ibid.*, pp. 148 *ss.*

perseguidor. Por el contrario, la presa utiliza la llamada "demostración proteica", cambiando continuamente de dirección, al azar, lo que confunde al perseguidor, sin depender del sendero de este último.[39] Según Curio, "relativamente pocos depredadores han superado la defensa proteica",[40] pero uno de éstos debe de ser el chita, citado por el propio Curio, que utiliza una velocidad inferior a la máxima al perseguir a su víctima porque la "desaceleración permite al felino seguir cada desviación de la impredecible fuga de su presa".[41]

Esto podría describirse con el argumento siguiente. En una primera etapa evolutiva, depredador y presa siguen la sencilla estrategia de la persecución y la evasión directas. En una segunda etapa, el depredador mejora la estrategia indirecta de intercepción en el camino. En una tercera etapa, la presa mejora la contraestrategia proteica, y en una última etapa, el depredador responde con el comportamiento de espera de la persecución a velocidad subóptima. Yo no sé si el verdadero proceso evolutivo fue en realidad así, pero sea como fuere, las etapas que lo integran están bien documentadas. El punto principal del penúltimo párrafo queda confirmado por la observación de que en esta secuencia, cada paso es, en realidad, ventajoso para la especie que lo da.

Esta digresión sobre la conducta del depredador y la presa muestra, en primer lugar, que los animales *sí* son capaces de una conducta maximizadora global y, en segundo lugar, que tal conducta se encuentra en situaciones sumamente específicas y estereotipadas.[42] Para nuestros propósitos

[39] Edmunds (1974), pp. 145-146; Curio (1976), pp. 153-154.
[40] Curio (1976), p. 153.
[41] *Ibid.*, p. 147; la cita fue ligeramente modificada.
[42] Barash (1977, pp. 3 *ss)* describe un experimento en el

actuales, la segunda idea es la básica. El rasgo característico del hombre no es una capacidad programada para utilizar estrategias indirectas o para adoptar la conducta de espera en situaciones específicas sino, antes bien una *capacidad generalizada de maximización global* que se aplica aun en situaciones cualitativamente nuevas. Dados ciertos hechos como la pleiotropía, la preadaptación y la deriva genética no es muy sorprendente, tal vez, que mutaciones de un solo paso pueden producir estrategias de dos pasos. Sin embargo, tales estrategias deben recibir una explicación separada en cada caso particular, y no hay un mecanismo general que pueda crear una presuposición de su presencia. Por contraste, la conducta globalmente maximizadora en el hombre queda explicada en forma inmediata por su capacidad de relacionar-

que se colocó a un perro y a una ardilla en una situación en que para obtener alimentos se necesitaba el uso de una estrategia indirecta, la cual logró la ardilla pero no el perro. Explica esto diciendo que en el caso de las ardillas, "la capacidad de hacer buenas desviaciones fue favorecida por la selección natural, de modo que cada población de ardillas llegó a estar compuesta de individuos que eran buenos para apartarse de las metas, con objeto de alcanzarlas". Hay que reconocer que el marco experimental descrito por Barash no es una "situación muy específica y estereotipada" en la vida de las ardillas. Sin embargo, parece ser lo bastante similar a las situaciones de la vida real en que la capacidad de "apartarse de las metas para alcanzarlas" se desarrolló, en forma que no constituyera una seria objeción a la tesis general de este ensayo. Hay aquí un continuo de casos, de modo que cuanto más generalizada sea la capacidad de estrategias indirectas, más nueva será la situación en que pueda ejercerse y más justificada la atribución de estados mentales e intenciones. En un extremo del continuo está la célula descrita por Rosen (n. 44 *infra*), y en otro extremo el "mono genio", descrito por Wilson (n. 43); las ardillas descritas por Barash se encontraban en un sitio intermedio.

se con el futuro y con lo simplemente posible. Puede elegir *la* alternativa globalmente mejor porque es capaz de analizar *todas* las alternativas, todos los futuros posibles. Aquí, antes bien, es la falta de maximización global la que necesita una explicación separada en cada caso. Tales incapacidades de alcanzar una maximización global son el tema del capítulo II.

Podemos decir que al *crear al hombre, la selección natural se ha trascendido a sí misma.* Este salto implica una transición de la adaptación no intencional, sea local o accidentalmente global, a la adaptación intencional y deliberada. El gran avance tal vez pueda observarse *in status nascendi* con "Imo, el mono genio" que, al dársele una mezcla de avena y de trigo, inventó la estrategia indirecta de arrojar, primero, la mezcla al agua (un paso atrás) y luego recoger los granos de trigo, más ligeros (dos pasos adelante). Edward O. Wilson, de quien hemos tomado esta descripción, califica esta conducta como "un elemento cualitativamente nuevo: arrojar por un momento la comida y aguardar un poco antes de recuperarla".[43] Yo diré que este elemento cualitativamente nuevo es, ni más ni menos, el despliegue de la capacidad generalizada de maximización global en una situación que, sin duda, era por entero nueva. Sugeriré, además, que con esta capacidad generalizada, la *mente* entra en la arena evolutiva. En los casos de estrategias indirectas, específicas de cada situación (o esperas) no necesitamos recurrir a estructuras intencionales o mentales. Algunos sin duda refutarán esto, en el caso de los ejemplos de la presa y el depredador, antes presentados, pero sea como fuere no puede haber ni huella de intencionalidad en la

[43] Wilson (1975), p. 171.

conducta de "una célula [que] puede, por decirlo así, 'ver' varias jugadas adelante, y hacer una jugada enérgicamente desfavorable si, al hacerlo, llega a una posición desde la que pueda hacer una extraordinariamente favorable, en una transición ulterior".[44] Por contraste, el empleo de estrategias globalmente maximizadoras en situaciones nuevas debe implicar un análisis del contexto, un examen de varias posibles jugadas y por último una lección deliberada entre ellas. "La sélection s'opère, non parmi les possibles, mais parmi les existants", escribió François Jacob[45] en una frase que se encuentra en el origen del enfoque actual. El hombre, una vez más por contraste, tiene la facultad de elegir entre posibles no realizados.

Yo creo que estas ideas pueden arrojar cierta luz sobre la cuestión recientemente tratada por Donald Griffin[46] y Daniel Dennett,[47] sobre si puede decirse —y en qué sentido— que los animales tienen una vida mental, intenciones, conciencia de ellos mismos, "imágenes mentales" de objetos ausentes, etc. Griffin, en cierto punto de su argumento, escribe: "Dado que, a la vez, sus congéneres y los observadores humanos pueden predecir la conducta futura de un animal a partir de sus movimientos de intención, parece notablemente imprudente suponer que el animal que ejecuta el movimiento de intención no puede prever el próximo paso de su propia conducta."[48] Esto, creo yo, es válido para los hombres, mas no para los animales. En la explicación de

[44] Rosen (1972), pp. 55-56.
[45] Jacob (1970), p. 313.
[46] Griffin (1976).
[47] Dennett (1976); *véase* también Dennet (1971).
[48] Griffin (1976), p. 44.

la conducta humana, la carga de la prueba la tiene el que niega la intencionalidad y la anticipación; en etología, ocurre al revés. Griffin, así mismo, tiene dificultades para que la idea de las intenciones animales resulte operacional: "Pero supongamos que un perro *sí* prefiera los acontecimientos que probablemente ocurrirán mañana. ¿Cómo podríamos reconocer este hecho, observando su conducta de hoy?" [49] Aquí, yo supongo que podríamos inventar pruebas, donde los animales tuviesen que hacer una elección entre una conducta localmente maximizadora y otra globalmente maximizadora, y tomar la aparición de esta última como prueba de una auténtica representación mental del futuro. Dennett, en su análisis de la "personidad", enumera seis normas que hay que satisfacer si se quiere que un agente cuente como persona. Aquí sólo ofreceré algunas observaciones sobre la relación entre sus normas tercera y cuarta. Según la tercera, una condición indispensable para que alguien sea una persona consiste en que lo tratemos como agente intencional. Dennett pasa entonces a observar que esto también incluye nuestra actitud hacia los animales y aun las plantas, y añade la norma (cuarta) de la *reciprocidad*. Es decir, una persona debe ser un agente "al que atribuimos no sólo creencias sencillas, deseos y otras intenciones, sino también creencias, deseos y otras intenciones *acerca* de creencias, deseos y otras intenciones".[50] En la terminología de la sección 4 una persona debe ser capaz de actuar sobre previsiones acerca de previsiones. Sin embargo, yo supongo que una norma menos rígida podrá prestar iguales servi-

[49] *Ibid.*, p. 50.
[50] Dennett (1976), p. 181.

cios; por ejemplo, la capacidad de maximización global.

4. La conducta estratégica en los animales y el hombre

En la sección 2 hicimos una breve referencia a una distinción, entre un medio paramétrico y un medio estratégico. (*Véase* también la sección 2 del capítulo III para esta distinción.) Afirmaré ahora que la selección natural siempre actúa como si el medio fuera paramétrico, aun cuando en realidad sea estratégico, mientras que el hombre es capaz de tomar en cuenta la naturaleza estratégica del contexto. (En la sección 2 traté de la relación existente entre una población y las otras poblaciones que integran su medio, mientras que aquí trataré de la relación existente entre un individuo y los demás individuos en la población a la que pertenecen. Sin embargo, el punto lógico es el mismo.) Mi principal argumento también puede plantearse diciendo que el hombre —en contraste con la selección natural— es capaz de concebir la solución de juegos en que ningún autor tiene una estrategia predominante.

El actor paramétricamente racional trata su medio como una constante, mientras que el actor estratégicamente racional toma en cuenta el hecho de que el medio está integrado por otros actores, y que él es parte de su medio, y que los demás lo saben, etcétera. En una comunidad de actores paramétricamente racionales, cada uno creerá que es el único cuya conducta es variable, y que todos los demás son parámetros para su problema de decisión. Hablando sobre estas creencias incongruentes, los actores generarán

consecuencias no intencionales y perversas del tipo analizado por los sociólogos, desde Marx hasta Keynes o desde Mandeville hasta Boudon.[51] Tal conducta es irracional desde un punto de vista colectivo, aun si individualmente cada actor satisface las condiciones de racionalidad. Una condición necesaria (pero insuficiente) para la racionalidad colectiva es la transición al pensamiento estratégico. En el modo de interacción estratégico, o las teorías y los juegos, cada actor debe tomar en cuenta las intenciones de todos los demás, incluyendo el hecho de que las intenciones de ellos se basan en sus expectativas concernientes a las suyas propias. Durante largo tiempo se creyó que esto implicaba una regresión infinita,[52] pero con frecuencia no es así. Utilizando el concepto de un *punto de equilibrio* es posible anular la regresión infinita y llegar a un punto de acción únicamente definible y predecible, que será elegido por hombres racionales. (Pero para excepciones, *véase* la sección 2 del capítulo III.) Desde nuestro actual punto de vista, el hecho decisivo es que los actores humanos no sólo toman sus decisiones sobre la base de sus expectativas del futuro, como ya se indicó al término de la sección 3, sino también *sobre la base de sus expectativas acerca de las expectativas de los demás*. Ésta es la auténtica mónada leibni-

[51] *Véase* Elster (1978a), capítulo V, para muchos ejemplos; *véase* también Boudon (1977).

[52] Rémont de Montmort (1713) escribió acerca de tal interdependencia mutua: "Les questions son très simples, mais je les crois insolubles. Si cela est, c'est un grand dommage car cette difficulté se rencontre en plusieurs choses de la vie civile. Quand deux personnes par exemple, ayant affaire ensemble, chacun veut se régler sur la conduite de l'autre." *Cf.* también la sección 2 del capítulo III, *infra*, para juegos que auténticamente carecen de una solución (no cooperativa).

ziana, en que cada individuo refleja la totalidad, desde su punto de vista. La transparencia y simetría de esta interacción hace que el destino de los actores esté en sus propias manos, mientras que una comunidad de actores paramétricamente racionales estará a merced de fuerzas causales que los eludirán y que frustrarán perpetuamente sus planes. Sea como fuere, los actores estratégicos lograrán escapar de lo que podríamos llamar fuerzas causales supraintencionales. En los capítulos II y III retornaremos a la sombría cuestión de la causalidad subintencional (en gran parte biológica).

Afirmaré ahora que no existe tal interacción estratégica en el reino animal, y que el no percibir esta diferencia entre los dominios social y biológico ha hecho que varios autores recientes se metan en un callejón sin salida. Para establecer la distinción en la forma más sencilla, consideraremos dos juegos que serán resueltos de manera diferente en una sociedad humana, pero de manera idéntica en las sociedades animales. El primero es el Dilema del preso que ha sido extensamente estudiado por psicólogos, economistas y politólogos.[53] Plantearé la estructura de este juego en forma un tanto insólita, en que aparece como un juego entre "yo" y "todos los demás". Las variables x y y cubrirán el conjunto de estrategias {C, D}, donde C representa la cooperación y D la defección, y $/x, y/$ significa "yo escojo x, todos los demás escogen y". El Dilema del preso queda definido, pues, por el hecho de que para cada actor (es decir, "yo") las alterna-

[53] Rappoport y Chammah (1965), Olson (1965), M. Taylor (1976a) son revisiones de aspectos psicológicos, políticos y filosóficos de este juego.

tivas pertinentes quedan en el siguiente orden de preferencia:

(1) /D, C/
(2) /C, C/
(3) /D, D/
(4) /C, D/

Inmediatamente puede verse que, hagan lo que hagan los demás, mi mejor estrategia será siempre D. Dado que los demás se encuentran en la misma posición, también ellos elegirán D, siendo el resultado perfectamente predecible /D, D/, lo que es peor *para todos*, que /C, C/. Esto muestra por qué la interacción estratégica no es condición suficiente de racionalidad colectiva. (Aquí, es la *suboptimalidad* del resultado la que destruye la racionalidad colectiva; en la sección 2 del capítulo III haremos un análisis de un problema más complejo.)

Entre los innumerables ejemplos de esta estructura que podemos tomar de las ciencias sociales podemos elegir el problema de la motivación en el trabajo. En una sociedad igualitaria, sin solidaridad entre los obreros, ninguno se esforzará en su trabajo, aun si sería mejor para todos que todos lo hicieran. Para un ejemplo biológico podemos recurrir a la explicación que da George Williams, de la evolución de los bancos de peces.[54] Si D representa la tendencia a situarse en el centro del banco, y C la ausencia de semejante tendencia, Williams arguyó que en realidad tenemos aquí un Dilema del preso en que la formación de bancos evolucionó porque es individualmente favorable aun si colectivamente puede conducir a la extinción de toda la población. *Ceteris paribus* siempre es mejor estar en el centro del banco

[54] G. C. Williams (1966), pp. 212 ss.

que en la parte exterior, porque los peces que están en el centro se encuentran menos expuestos a los depredadores. Sin embargo, si cada quien trata de colocarse en el centro, el banco, en conjunto, se vuelve más compacto y más expuesto a sus depredadores. La agresión nos ofrece un ejemplo más sencillo: en una población de organismos pacíficos, un mutante agresivo predominará, pero sus descendientes podrán atacarse unos a otros, hasta llegar al exterminio.

El segundo juego, que compararemos con el Dilema del preso es menos conocido. Fue extensamente estudiado por Amartya K. Sen, bajo el rubro de "Juego del seguro",[55] y difiere del Dilema del preso en el orden de las dos primeras alternativas:

(1) /C, C/
(2) /D, C/
(3) /D, D/
(4) /C, D/

Este juego difiere del Dilema del preso en el rasgo fundamental de que *no hay estrategia dominante*, es decir, ninguna estrategia es mejor, hagan lo que hagan los demás. ¿Cómo debe comportarse un actor racional en este juego? Si no tiene una información perfecta, deberá utilizar la norma maximin y elegir la estrategia que le garantice el rendimiento mínimo más alto. En el Juego del seguro esta es la estrategia D, pues lo peor que puede ocurrirle con esta elección es la alternativa (3), mientras que lo peor que podría ocurrirle si eligiera C es la alternativa (4). Por otra parte, si suponemos que todos los jugadores disponen de información perfecta, entonces todos ellos elegirán C, en la confianza de que to-

[55] Sen (1967, 1973, 1974).

dos los demás harán lo mismo, ya que nadie tiene nada que ganar si actúa de otra manera. En el Dilema del preso el requerimiento de plena información no es necesario para que surja la solución, pero en el Juego del seguro, la menor incertidumbre o desconfianza hará que un actor elija D en lugar de C. Sólo en una situación perfectamente transparente convergirán los actores en la conducta colectivamente racional. Dada una información perfecta, puede expresarse la diferencia entre los dos juegos diciendo que en el Dilema del preso el óptimo /C, C/ es, a la vez, individualmente accesible (nadie dará el primer paso hacia él) e individualmente inestable (todos darán el primer paso, apartándose de él) mientras que en el Juego del seguro, es individualmente inaccesible e individualmente estable. *(Véase* la sección 2 del capítulo II para el juego de "la gallina", donde el óptimo es individualmente accesible e individualmente inestable.) Dado que los actores humanos pueden actuar sobre sus expectativas acerca de expectativas, la estabilidad del óptimo en el Juego del seguro les hará converger hacia él a pesar de la inaccesibilidad individual.

Para un ejemplo sociológico del Juego del seguro, podemos contemplar de nuevo el problema de la motivación en el trabajo, en que Sen ha argüido que en una sociedad socialista, las preferencias debieran reflejar el Juego del seguro, y no el Dilema del preso. En el socialismo, el esfuerzo del trabajo arduo no debiera ser (al menos eso espero) una estrategia dominante. Nadie preferiría un trabajo arduo, muchas veces monótono, por el trabajo mismo, si tomara en cuenta lo que los demás estuvieran haciendo. Sin embargo, la *solidaridad* con los demás podría pro-

ducir el esfuerzo requerido. Si estoy seguro de que los demás ponen de su parte en el trabajo, bien puedo preferir hacer el mío que convertirme en un "lobo solitario". La solidaridad es *altruismo condicional,* a diferencia del altruismo incondicional del imperativo categórico y del egoísmo incondicional de la sociedad capitalista. (*Cf.* también la sección 7 del capítulo III, para otras observaciones sobre el altruismo.)

Es un problema empírico determinar si el Juego del seguro a menudo se resuelve, en el sentido de que /C, C/ surge como resultado de cálculos individuales. Puede suponerse que los muy parcos requerimientos de información sólo pueden satisfacerse si la información llega del exterior (por ejemplo, del gobierno, actuando como coordinador) o si la comunidad es tan pequeña y estable que cada quien puede llegar a conocer en realidad a todos los demás. Sin embargo, en nuestro marco actual, nos importa menos la *probabilidad* de una conducta cooperativa que la *probabilidad* misma de una acción cooperativa mutuamente dependiente. En las sociedades humanas, puede realizarse esta posibilidad; no así en la selección natural. Afirmaré ahora que una falacia, bastante común, de quienes escriben sobre la selección natural consiste en pensar que la *estabilidad explica la evolución,* de modo que el surgimiento de la evolución en un juego sin estrategias dominantes puede explicarse por su inmunidad ante los tipos aislados.[56] Se requiere acce-

[56] A esta conclusión llegó independientemente Michael Taylor (1976b), quien escribe: "El problema... es que, aunque resulte razonable suponer que un equilibrio, una vez alcanzado, se mantendrá, no hay razones para suponer que será alcanzado en principio por quienes se encuentren

sibilidad individual para que se llegue a la solución en los juegos biológicos, y la estabilidad individual sólo puede asegurar su mantenimiento una vez que ha evolucionado. Procederé ahora a examinar algunos ejemplos de esta falacia.

En el libro de E. O. Wilson, *Sociobiology* —por lo demás, admirable—, encontramos un primer ejemplo de razonamiento falaz, pasando de estabilidad a surgimiento. En su capítulo "Los primeros motores de la evolución social", expresión que definitivamente sugiere ciertos mecanismos para el *surgimiento*, y no sólo para el mantenimiento de la conducta, encontramos la siguiente explicación de los bancos de peces:

> Una variación potencial de la estrategia de la horda egoísta es la utilización de un "protector" que consume parte de la población pero que lo compensa, con creces, excluyendo a otros depredadores. El muy común pez de los corales *Pempheris oualensis* forma bancos de pocos cientos o miles de individuos que se refugian durante el día en unos agujeros bien sombreados, pasajes de coral y huevas, de frente al mar abierto. Comparten estos refugios con uno o con unas cuantas clases de peces depredadores, casi siempre el cerránido *Cephalopholis argus*, que se alimenta de ellos en cantidades limitadas. Como los depredadores son territoriales, el *Pempheris* sale ganando, hasta cierto punto, al andar en bancos y así limitar su exposición durante las horas del día a sólo uno o unos cuantos de sus enemigos. Al saturar en conjunto a los depredadores favorecidos con más de lo que puedan consumir, los miembros individuales del banco se ven favorecidos con una mayor probabilidad de supervivencia.[57]

en el caso de un juego al que jueguen agentes humanos racionales."

[57] Wilson (1975), p. 38.

Según yo interpreto este texto, *afirma* que el compartir escondites con un depredador es una adaptación individualmente estable, e *implica* que no se necesita ninguna otra explicación del surgimiento de este comportamiento. La adaptación es estable porque cualquier pez en particular de quien tratara de practicarla por su cuenta en el mar abierto se encontraría más expuesto a la depredación que en las cavernas. Por otra parte, esto no prueba que la adaptación sea individualmente accesible (por ejemplo, una estrategia dominante); en realidad, la "seguridad en los números" por medio de la saturación conjunta de los depredadores no pudo favorecer al pez que por primera vez se aventuró en las cavernas. Debe de haber algún umbral más allá del cual se encuentra la seguridad en los números (como lo reconoce el propio Wilson en otro contexto)[58] y la evolución hasta llegar a este umbral no puede explicarse invocando la causalidad retroactiva o las previsiones. Una ambigüedad similar se encuentra en el análisis que hace Curio de "satisfacer el apetito de los depredadores" como contraestrategia general de la presa.[59]

Un segundo ejemplo se refiere al uso de *estrategias mixtas* en juegos animales y humanos. Una estrategia mixta es una distribución de probabilidades, fijando para cada estrategia pura la probabilidad de que sea escogida. No hay que confundir una población de actores que se valgan de estrategias mixtas con una población mixta, ni con una población de actores cada uno de los cuales cubre *físicamente* (y no sólo probabilísticamente) varias estrategias.[60] El surgimiento de

[58] *Ibid.*, p. 212.
[59] Curio (1976), pp. 33-34.
[60] Así, en la carrera de caballos, la misma macrodis-

la teoría de los juegos como disciplina se basa en la prueba de Von Neumann, de 1928, de que algunos juegos aparentemente sin solución sí admiten una solución con estrategias mixtas. Ahora bien, una auténtica estrategia mixta, que incluya probabilidades distintas de cero al menos para dos estrategias puras, nunca puede ser una estrategia predominante. Para los juegos sociales, esto implica que la realización de la solución requiere información perfecta, y para los juegos animales, implica que la solución sólo puede realizarse por accidente. Exploremos estas dos proposiciones por medio de unos ejemplos.

Entre los escasos intentos de explicar los problemas de la vida real (y no militares) mediante la noción de estrategias utilizadas, podemos citar la reciente obra de Raymond Boudon sobre educación y privación relativa.[61] En su obra sobre el último problema, Boudon trata de explicar el descubrimiento ya clásico y paradójico de *The American Soldier*, de que el nivel de frustración en el ejército era el más alto donde mayores eran las posibilidades objetivas de frustración. Imaginemos, en bien de la sencillez, una cohorte de 20 policías militares y otra cohorte de 20 aviadores. Los primeros pertenecen a un grupo carac-

tribución de apuestas puede ser generada por tres mecanismos distintos: cada persona apuesta todo su dinero a un caballo según cierta distribución de probabilidades (una estrategia mixta propiamente dicha); cada persona distribuye su dinero sobre muchos caballos (una mezcla física) y cada persona apuesta todo su dinero deterministamente a un caballo; diferentes individuos escogen diferentes caballos. Para la relación entre el primero y el segundo concepto, *véase* Dorfman *et al.* (1958), pp. 400-401, citando de J. D. Williams (1954), p. 103. Para la relación entre el primero y el tercero, *véase* el análisis de Maynard Smith y de Dawkins, *infra*.

[61] Boudon (1977), Caps. IV-V.

terizado por escasas posibilidades de ascenso y un bajo nivel de frustración, mientras que entre los segundos, tanto el potencial de ascenso como la frustración son grandes. Supongamos, además, que para tener alguna posibilidad en el juego de los ascensos, hay que hacer un esfuerzo extra equivalente a un franco. El ascendido recibe el equivalente de cinco francos en remuneración. Cinco de 20 pilotos y dos de cada 20 policías militares lo lograrán, si todos los participantes (es decir, los que invirtieron un franco) tienen la misma posibilidad de éxito. Entonces, es fácil ver que los 20 aviadores invertirán un franco, resultando decepcionados 15 de los 20, mientras que sólo 10 policías militares entrarán en el juego, 8 de los cuales quedarán decepcionados. La razón de que sólo 10 policías militares participen en el juego es que cada uno de ellos considera individualmente racional utilizar la estrategia mixta de entrar en el juego con una probabilidad de ½. Esta demostración, ingeniosa y elegante, adolece del defecto de que el empleo de la estrategia mixta sólo es racional si cada persona está segura de que las demás harán lo mismo, lo que requiere una inverosímil cantidad de información acerca de los otros jugadores. (Para una dificultad adicional, *véase* también la sección 2 del capítulo III.) Sin embargo, este defecto es sólo empírico, y no hay nada intrínsecamente imposible en el modelo. *Si los seres humanos fuesen siempre tan racionales como pueden serlo a veces, así es como se comportarían.*

No puede darse semejante justificación general al surgimiento de la solución en los juegos animales que requieren estrategias mixtas. En realidad, a este respecto, las estrategias mixtas son aún peores que las estrategias puras no dominan-

tes. Las últimas pueden surgir como soluciones a un juego en que *son* estrategias dominantes, y luego sostenerse aun cuando el juego se transforme (por medio del cambio ambiental) en un juego en que son individualmente estables sin ser individualmente accesibles. Esto no puede ocurrir con las estrategias mixtas, porque no pueden ser predominantes en ningún juego. (Por supuesto, pueden quedar fijas por la deriva genética.) Para encontrar un ejemplo en que se emplean estrategias mixtas en la teoría evolutiva, podemos echar una mirada a la "lógica del conflicto animal", tal como fue analizada por John Maynard Smith [62] (o, antes bien, la versión simplificada de su análisis, presentada por Richard Dawkins).[63] Aquí, suponemos que la población consiste en dos genotipos, halcones y palomas; éstas no son las especies biológicas, sino las dos variantes conductistas de una misma especie. Cuando una paloma encuentra a otra paloma, cada una de ellas tiene 50 % de oportunidad de ganar 50 "unidades evolutivas" y 50 % de perder 10 unidades. Cuando un halcón encuentra a un halcón, cada uno tiene 50 % de oportunidad de ganar 50 y 50 % de perder 100. Cuando un halcón encuentra a una paloma, el primero gana 50 y la segunda pierde 10 con toda certidumbre. Entonces, es fácil mostrar que la proporción de palomas a halcones en la población convergirá hacia 5:7, es decir, un polimorfismo estable. Hasta allí, todo va bien. Luego viene el paso decisivo e ilegítimo, cuando a la vez Maynard Smith y Dawkins afirman que una situación similar puede evolucionar sin polimorfismo, a saber, si cada individuo emplea la estrategia de la paloma con

[62] Maynard Smith (1973, 1974).
[63] Dawkins (1976), Cap. v.

una probabilidad de 5:12 y la estrategia del halcón con probabilidad de 7:12. Esto sería cierto, en realidad, en una población de actores humanos que dispusieran de información perfecta, pero no lo es para actores animales incapaces de crear una estrategia que sólo es óptima contra la misma estrategia elegida por todos los demás. A la inversa, el polimorfismo (es decir, la población mixta a la que antes nos referimos) nunca podría surgir en una población humana porque no hay manera de decidir tácitamente cuáles individuos adoptarán cuál estrategia.

Para un tercer ejemplo de la falacia podemos ver un profundo artículo de Gary Becker,[64] quien arguye que las nociones de egoísmo y de altruísmo son más sutiles y entrelazadas de lo que habitualmente se ha creído. En términos generales, el argumento dice que un egoísta tiene un claro incentivo de actuar como altruista, porque un aumento del ingreso del altruista rebotará sobre el egoísta. Aquí, desde luego, una suposición fundamental es que el egoísta es capaz de prever correctamente la acción del altruista. Renunciar a un aumento de ingreso que es producido a expensas de una reducción aún mayor de ingreso para el altruista no puede ser una estrategia dominante del egoísta. Sólo es eficaz contra altruistas y no contra otros egoístas. Entre las sorprendentes conclusiones del análisis de Becker se encuentra la de que un altruista puede estar en realidad mejor que un egoísta, porque los "benéficos efectos indirectos sobre la conducta de los demás pueden dominar las 'desventajas' directas de ser altruista".[65] Becker aplica entonces el mismo esquema al problema de la aptitud genética,

[64] Becker (1976), Cap. XIII.
[65] *Ibid.*, p. 287.

donde afirma que recurrir a la selección de grupo o la selección de parentesco para explicar la conducta altruista es menos indispensable de lo que casi todos los biólogos lo han creído. Si los genes altruistas confieren beneficios individuales a sus poseedores porque otros los toman en cuenta, para maximizar *sus* beneficios, entonces ordinariamente la selección natural al nivel individual basta para explicar el surgimiento del altruismo.

En la sección 4 del capítulo II y la sección 7 del capítulo III volveré a algunos aspectos de la parte sociológica de este argumento. Aquí deseo cuestionar su parte biológica. Consideremos una población de egoístas en que aparece por primera vez un mutante altruista. Los egoístas, desde luego, lo tratarán como si fuera sólo otro egoísta, porque no tienen modo de "saber" o de "prever" que el altruista les dará algunos de los beneficios que ellos le den a él. Por esta razón, el gene altruista sólo tendrá pérdidas y no ganancias, y pronto desaparecerá de la población. En otras palabras, Becker parece atribuir a organismos inferiores la capacidad de actuar de acuerdo con expectativas que, según mi argumento, son específicas del hombre. Y aun si los organismos tuviesen esta capacidad, yo sugeriré (tentativamente) que, antes bien, razonarían de esta manera: "¿Por qué hemos de corresponder al altruista, ya que esto nos dejaría peor, comparados con él, aun si en términos absolutos mejoraríamos nuestra suerte?" Lo que queda maximizado en la selección natural es la aptitud diferencial, no la aptitud absoluta. Como lo notó Alexander Gerschenkron en otro juego de "lo que se gana, se pierde", "es indeseable un tráfico mutuamente benéfico que también enriquece a los enemigos".[66]

[66] Gerschenkron (1970), p. 65.

Como última observación, deseo hacer ver que aun en el reino animal, *pueden* surgir soluciones de teoría de los juegos en ausencia de estrategias dominantes, por ejemplo en el caso en que las sucesivas adaptaciones a las adaptaciones de otros convergen hacia la solución (como en un ciclo de la telaraña húmeda). Un ejemplo de este caso nos lo ofrece W. D. Hamilton en su análisis de teoría de los juegos, de las "extraordinarias proporciones de los sexos".[67] Estudió el caso del parasitismo doble (dos parasitoides en un anfitrión); más precisamente, la situación en que hay dos tipos de hembras cada una de las cuales está tratando de maximizar su aptitud genética mediante la selección de una proporción de sexo (la proporción de machos para la progenie total) que es óptima relativa a la proporción elegida por la otra. Arguyendo, para empezar, que la aptitud es proporcional al número de inseminaciones por los hijos más el número de hijas, Hamilton muestra que para una hembra que está utilizando la proporción de sexo x en un anfitrión, con otra hembra que está utilizando la proporción x_0, la aptitud será proporcional a

$$\frac{x}{x + x_0}((1 - x) + (1 - x_0)) + (1 - x)$$

Para la hembra que utilice x_0, la aptitud será proporcional a la misma expresión, con x y x_0 cambiando de lugar por doquier. Si ahora limitamos el conjunto de posibles proporciones a $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ —una simplificación que no establece ninguna diferencia para el análisis— surge la siguiente matriz de rendimientos.

[67] Hamilton (1967).

		x_0				
		0	1/4	1/2	3/4	1
x	0	?	1, 2.5	1, 2	1, 1.15	1, 1
	1/4	2.5, 1	1.5, 1.5	1.2, 1.3	1, 1	0.9, 0.6
	1/2	2, 1	1.3, 1.2	1, 1	0.8, 0.7	0.7, 0.3
	3/4	1.5, 1	1, 1	0.7, 0.8	0.5, 0.5	0.4, 0.1
	1	1, 1	0.6, 0.9	0.3, 0.7	0.1, 0.4	0, 0

Sin tomar en cuenta el caso indeterminado (y biológicamente trivial) de que ambas hembras decidieran no producir machos, inmediatamente vemos que el único punto de equilibrio (y por ello, la solución) de la matriz es cuando ambas jugadoras emplean ¼. También observamos que la proporción ¼ *no* es una estrategia dominante, pues si x escoge 1, x_0 escogerá 0 en lugar de ¼, y a la inversa. Sin embargo, Hamilton está justificado cuando dice que "por medio de prueba y error, dos jugadores ingenuos pronto aprenderían que jugar constantemente a ¼ era la estrategia de rendimiento óptimo".[68] Por ello, a saber, si x_0 comienza con uno del grupo $\{0, ¼, ½, ¾\}$, entonces ¼ es la mejor respuesta de x, y en la siguiente ronda, ¼ también será la mejor respuesta de x_0. Si x_0 empieza con 1, x contraatacará con 0 a lo cual x_0 replicará con ¼, lo que a su vez inducirá a x a escoger ¼. Así, cualquiera que sea la estrategia inicial escogida por x_0, la secuencia de respuestas y contrarrespuestas hará que las jugadoras converjan en la solución, en un máximo de cuatro pasos. Esto muestra que la presencia de una estrategia dominante no es condición necesaria para que surja la solución a la teoría

[68] *Ibid.*, p. 486.

de los juegos, pero, a pesar de todo, sigue siendo válido que a falta de pruebas específicas en sentido contrario, no puede presuponerse que la solución se encontrará en un juego sin estrategias dominantes. Hamilton ofrece tal argumento; no así Wilson, Maynard Smith y Becker. El procedimiento de éstos parece basarse en suposiciones injustificadas concernientes a la responsabilidad de la prueba; dan por sentado lo que queda por demostrar.

5. La explicación funcionalista en sociología

En esta parte, enfocaré cuestiones más familiares, que sin embargo, parecen dignas de reconsideración. Argüiré aquí que la explicación funcionalista en sociología, se basa en una mal concebida analogía tomada de la biología, y que un análisis más minucioso de las supuestas explicaciones funcionalistas mostrará que virtualmente en todos los casos falta uno o más de los rasgos definitorios. Ahora bien, no hay un consenso general sobre qué rasgos debieran entrar en la definición de una explicación funcional, y al lector cuya definición difiera de la mía, lo que sigue le parecerá improcedente. Y sin embargo, yo creo que mi empleo del término está en bastante buen acuerdo con las exposiciones de Robert Merton y de Arthur Stinchcombe,[69] por lo que no estoy levantando un espantapájaros para que lo derriben.

Así pues, de acuerdo con mi definición, *una institución o una pauta de conducta X queda explicada por su función Y para el grupo Z* si y sólo si:

[69] R. K. Merton (1957); Stinchcombe (1968).

1) Y es un *efecto* de X;
2) Y es *benéfica* a Z;
3) Y no fue *intencional* por los actores que produjeron X;
4) Y (o al menos la relación causal entre X y Y) *no es reconocida* por los actores de Z;
5) Y mantiene a X por un círculo de *retroalimentación* causal que pasa por Z.

En este paradigma, los actores que participan en X pueden ser idénticos con el grupo Z, o la relación puede ser de inclusión, traslape o separación. Proceden aquí algunos comentarios sobre las normas. Creo yo que es casi indiscutible que las funciones son una subespecie de los efectos. Decir que una institución "tiene una función" debe implicar que *hace* algo, es decir, genera algún efecto. La noción de un efecto benéfico es un tanto más oscura. Consideraremos que significa que Y es un máximo local de algún estado variable, del que los actores de Z siempre desean más, y no menos. Sería indebidamente restrictivo exigir que Y fuese un máximo local, pues entonces el paradigma biológico se desplomaría desde el principio. Huelga decir que los actores que producen X pueden estar dedicados a una actividad de maximización global (de alguna otra variable). El hecho de que Y no sea intencional ni reconocida es parte integral del paradigma mertoniano de las funciones latentes, al que estoy limitándome aquí. Por otra parte, es difícil, pero no imposible, descubrir la quinta norma en la versión que da Merton del funcionalismo.[70] Sin embargo, en el análisis de Stinch-

[70] En el paradigma explícito Merton (1957, pp. 50 *ss)* enumera once dimensiones de su análisis; el círculo de retroalimentación no está entre ellas. Sin embargo, en su análisis más informal, Merton implica varias veces que

combe, la existencia de un círculo de retroalimentación causal se convierte en rasgo central de la explicación funcional, y yo creo que esto también refleja la opinión de casi todos los que escriben sobre la teoría funcionalista.

Argüiré ahora toda una serie de proposiciones. Primero, es casi imposible encontrar algunos casos de análisis funcional en sociología, donde esté demostrada la presencia de todos los rasgos *1)-5)*. En segundo lugar, existe una ingenua clase de análisis funcional (criticado por Merton) que, a partir de la presencia de los rasgos *1)*, *3)* y *4)* incluye la presencia del rasgo *2)* y a menudo del rasgo *5)*. En tercer lugar existe una clase más refinada de funcionalismo (representada por Merton) que a partir de la presencia de los rasgos *1)-4)* concluye falazmente la presencia del rasgo *5)*. En cuarto lugar, hay una clase de funcionalismo (especialmente marxista) que arguye —falazmente— que una institución puede ser mantenida por sus efectos a *largo plazo*, si éstos satisfacen las normas *2)-4)*, aun cuando los efectos a corto plazo sean dañinos y no benéficos. En quinto lugar —volviéndonos ahora a observaciones más positivas— los análisis que satisfacen las normas *1)-3)* y la norma *4)* o la norma *5)*

la adscripción de funciones (latentes) a alguna pauta explica la existencia y la persistencia de la pauta. Esta es, en realidad, la idea central de las páginas (64 *ss*) en que explica que la distinción entre funciones manifiestas y funciones latentes "aclara el análisis de pautas aparentemente irracionales". La afirmación más explícita (con una rara aparición del término "explica") aparece en una reinterpretación de Veblen: "Entre estas funciones latentes, que ayudan a explicar la persistencia y la ubicación social de la pauta de consumo ostensible, está su simbolización de 'fuerza pecuniaria y, así, de conquistar o retener un buen nombre'" (p. 69).

son de importancia fundamental en las ciencias sociales.

Empezando con la última proposición, las versiones que satisfacen los requerimientos *1), 2), 3)* y *5)* pueden ser llamadas *explicaciones de filtro*, mientras que los análisis que satisfacen las normas *1)–4)* pueden ser llamadas *explicaciones de mano invisible*.[71] La investigación hecha con dinero militar puede ser analizada por una explicación de filtro. Si personal académico solicita fondos de los militares para poder efectuar la investigación que de todos modos habría hecho (es decir, con dinero de cualquier otra fuente), el Departamento de la Defensa puede servir como filtro que selecciona algunas solicitudes y rechaza otras.[72] La composición resultante de la investigación será benéfica a los intereses militares, aunque esto no se lo hayan propuesto en absoluto los científicos individuales, quienes podrán decir, con verdad, que nadie les dijo lo que debían hacer.[73] A esto también se le puede decir un *caso de selección artificial*, en que el círculo de retro-

[71] Estoy tomando esta terminología de Nozick (1974, pp. 18 *ss*), aunque la empleo para expresar una distinción que no coincide completamente con la suya.

[72] Glantz y Albers (1974).

[73] En un libro de memorias, por lo demás muy bello, Ulam (1976, p. 232) parece caer en esta trampa: "Aquellos eran los días de los contratos de investigación de la defensa. Incluso los matemáticos los recibían con frecuencia. Johnny [von Neumann] y yo comentamos cómo en algunas de las propuestas, los propios científicos apenas describían cuán útil sería su investigación para el interés nacional, mientras que en realidad eran motivados por la curiosidad científica de buena fe y un afán de escribir unos cuantos ensayos." El "mientras que" está mal, porque (o en la medida en que) el Departamento de la Defensa pudo examinar las propuestas y retener aquellas que le pareciera que tenían posibilidades de uso militar.

alimentación opera mediante los efectos *reconocidos* de la estructura cuya persistencia habrá que explicar (*véase* también la sección 4 del capítulo III para otro ejemplo).

Para encontrar algunos casos de explicaciones de mano invisible, también podemos dirigirnos a la sociología de la ciencia, por ejemplo, a la propia obra de Merton, sobre descubrimientos múltiples y sobre el "efecto Matthew" en la ciencia.[74] En el primer ejemplo, tenemos que *X* es igual a los descubrimientos simultáneos, *Y* a una cantidad funcionalmente óptima de redundancia, y *Z* a la comunidad científica. En el segundo ejemplo, tenemos a *X*, igual a una cantidad desproporcionada de reconocimiento a los científicos consagrados, en caso de autoría múltiple y descubrimientos simultáneos, *Y* es la más rápida diseminación de resultados importantes, y *Z* la comunidad científica. En ambos casos, Merton demuestra, en forma convincente, la presencia de los rasgos *1)-4)* del paradigma funcionalista. También, en ambos casos, se vale del término "función", que *puede* ser interpetrado como afirmación implícita de que también se ha satisfecho la norma *5)*. Como ya lo observamos, Merton se muestra un tanto ambiguo en este punto. Personalmente, yo tengo la impresión de que Merton cree que el análisis funcional aporta también una *explicación* de los fenómenos a los que se imputan estas funciones. Como ejemplo último y más complejo del análisis de mano invisible, podemos tomar el caso en que *X* es igual a los inventos que ahorran mano de obra, *Y* es la reducción del nivel salarial, y *Z* es la comunidad de empresarios capitalistas.[75] Aquí, los actores que

[74] Merton (1973), cap. XX.

[75] Para más discusión acerca de este problema, *véase*

participan en *X* son los mismos que en el grupo *Z*, lo cual significa que el análisis *intencional* y el análisis *funcional* surgen como dos enfoques alternos (e igualmente falaces).[76]

Yo sólo conozco un ejemplo en que en realidad están presentes todos los rasgos *1)-5)*, o sea, en el intento hecho por la escuela de economistas de Chicago por explicar la maximización de las ganancias como resultado de la "selección natural" de las firmas hecha por el mercado.[77] La anomalía que condujo a este intento fue la siguiente. Por una parte, la conducta externa observada (elección de combinaciones de factor y nivel de producción) pareció indicar que las empresas en general adoptan una conducta de maximizar las ganancias, adaptándose a las condiciones del mercado. Por otra parte, el proceso interno de toma de decisiones de la empresa no pareció guiado por este objetivo; antes bien, parecieron típicas ciertas reglas aproximadas, ya establecidas. Para colmar la brecha entre la producción de la caja negra y su funcionamiento interno, los economistas en cuestión postularon que algunas firmas simplemente usan ciertas reglas generales de maximización de la ganancia, y otras no; que las primeras sobreviven mientras que las segundas se extinguen; que la rutina que tiende a maximizar la ganancia cunde por la población de empresas, sea por imitación, sea por adquisición. Entonces, si consideramos a *X* igual a

Elster (1978a), capítulo V y la sección 2 del capítulo III, *infra*; *cf.* también las referencias en n. 27, *supra*.

[76] Es decir, cuando todos se comportan de un modo que va en su beneficio como grupo, resulta tentador concluir *o bien* que actúan así para obtener esos beneficios, *o bien* que su conducta tiene esta función.

[77] *Cf.* la sección 4 del capítulo II y la sección 5 del capítulo III, *infra*.

cierta regla general, *Y* a la maximización de ganancias y *Z* al conjunto de las firmas, tendremos un paradigma de análisis funcional. El argumento está claramente modelado sobre una analogía biológica, y funciona (en la medida en que funcione)[78] sólo porque las nociones de aptitud, supervivencia, reproducción y herencia pueden transferirse sin excesiva modificación.

En este caso, el círculo de retroalimentación queda explícitamente demostrado. Sin embargo, las más de las veces es más postulado que demostrado. Un numeroso cuerpo de bibliografía sociológica parece basarse en una idea reguladora implícita de que si podemos demostrar que una pauta determinada tiene efectos no intencionales, no reconocidos y benéficos, entonces hemos explicado por qué existe y persiste. Creo yo que si el lector abre al azar cualquier libro sobre desviación, delincuencia o conflicto encontrará afirmaciones que apoyan mi interpretación. Una de tales interpretaciones, realmente tomada al azar de mi estante, podría ser la siguiente: "El conflicto dentro de estructuras burocráticas y entre ellas aporta los medios para evitar la osificación y el ritualismo que amenazan su forma o su organización."[79] Al utilizar la palabra "medios" el autor se compromete implícitamente con una explicación intencional o funcional del conflicto. Por otra parte, sus pruebas sólo apoyan la afirmación de que el conflicto puede tener el *efecto*

[78] Además de los problemas analizados en la sección 5 del capítulo III podemos añadir la dificultad local-global: a menos que supongamos que todas las posibles reglas generales están representadas en la población de empresas, no hay ninguna suposición de que este mecanismo permita alcanzar un máximo global.

[79] Coser (1971), p. 60. Del mismo volumen de lectura, Himes (1971) está en la misma vena.

de prevenir la osificación. No se delinea ningún círculo de retroalimentación (y desde luego, no se demuestran ningunas intenciones); sin embargo, el total descuido de la cuestión de las *causas* de los conflictos sólo puede explicarse por la suposición de que el autor postula en forma tácita una relación de retroalimentación por la que el efecto mantiene su causa. Mi manera de dar sentido a este enfoque consiste en imputar a estos autores la idea reguladora de que las normas *1)–4)* nos permiten inferir la satisfacción de la norma *5)*, a falta de pruebas en sentido contrario.

Arthur Stinchcombe se encuentra entre los relativamente pocos escritores que han reconocido que "la naturaleza de los vínculos causales inversos, a partir de las consecuencias de retraso a la estructura (es) un punto débil de la teoría funcional".[80] También esboza cinco mecanismos distintos que podrían colmar esta brecha. El primero es la selección natural en el sentido natural (biológico). El segundo y (creo yo) el sexto invocan la "selección natural" de las firmas por el mercado. El tercero se basa en la planificación consciente, y en realidad está fuera de lugar en este contexto. El quinto parece ser muy similar a lo que antes llamamos "mecanismo de filtración". El cuarto se explica en los siguientes términos enigmáticos: "Sin planificación, la gente puede encontrar satisfactorias las secuencias de la conducta. Así pueden mantenerse los servicios eclesiásticos sin mucha planeación, para alcanzar sus fines teológicos, porque la gente encuentra satisfactoria la interacción social, o la respetabilidad."[81] Este ejemplo durkheimiano ha sido analizado con cierto detalle por I. C. Jarvie, quien

[80] Stinchcombe (1968), p. 85.
[81] *Ibid.*, p. 86.

concluye que el análisis funcionalista no tiene poder explicativo, porque las "razones [de la gente] para ir a la iglesia pueden tener muy poco que ver con lo que están haciendo al ir a la iglesia".[82] En general, yo estoy de acuerdo con esta conclusión (que es apoyada por un argumento que el lector interesado deberá consultar), lo que significa que el único buen ejemplo de análisis funcional en la sociología es el que ya he señalado, o sea, la selección de empresas por el mercado.

Para comprender el origen de la sociología funcionalista podremos empezar observando que en biología son válidas las dos siguientes ideas reguladoras:

— Cada estructura persistente tiene una función (es decir, tiene efectos benéficos por medio de las de los que se mantiene).
— Cada estructura persistente, con efectos benéficos, tiene una función (más específicamente, se mantiene por medio de estos ejemplos).

Si estas proposiciones fueran interpretadas como leyes generales, la segunda sería redundante, pero la lógica de las ideas reguladoras es distinta. Una idea reguladora admite repeticiones, y es más fuerte cuanto menos sean los contraejemplos. Ya que la clase de contraejemplos a la segunda proposición es más pequeña que la clase de contraejemplos a la primera,[83] la segunda merece plan-

[82] Jarvie (1968), pp. 198-199.

[83] Se encuentran contraejemplos a la primera proposición en cuatro niveles distintos. Al nivel más fundamental, no debiéramos pensar que las mutaciones "tienen la función" de generar el proceso evolutivo. La selección natural no es resultado de la selección natural (G. C. Williams, 1966, p. 128); las mutaciones requieren explicaciones causales, no funcionales. En el nivel intermedio,

tearse por derecho propio. En el anterior análisis de Merton y Stinchcombe concluye diciendo que en las ciencias sociales no puede haber idea reguladora análoga a la segunda de estas proposiciones. No puede haber una suposición general —la carga de la prueba recae sobre quien sostiene la opinión opuesta— de que la subclase pertinente de efectos benéficos mantiene sus causas. *A fortiori* esto debe valer para la analogía sociológica de la primera idea reguladora, que es justamente criticada por Merton.[84] Por tanto, yo creo que ha llegado el momento de que la sociología se sacuda por completo su herencia biológica. Merton prestó un servicio importante al criticar la idea reguladora de que "todo tiene una función", correspondiente a la primera de las dos proposiciones anteriores. También efectuó brillantes análisis de la variedad de mano invisible, pero fue llevado a exagerar los efectos y a desdeñar las causas de las pautas de conducta mediante alguna suposición implícita, similar a la segunda proposición.

Terminaré con algunas breves observaciones sobre marxismo y funcionalismo. Un rasgo extraño y triste de la sociología marxista contemporánea y de la ciencia política es la adopción de los

la teoría de la evolución no darwiniana arguye que el número de sitios sin función (o antes bien, ocupantes de sitios) en las proteínas es muy grande en realidad. Al nivel más molar, la pleiotropía es un mecanismo general para generar y mantener rasgos neutrales o hasta dañinos. Los contraejemplos del cuarto nivel también son contraejemplos a la segunda proposición: se refieren a los efectos benéficos que brotan en la suma estadística de muchas adaptaciones individuales. G. C. Williams (1966, p. 209) se refiere a la errónea concepción "de que cuando alguien demuestra que cierto proceso biológico produce un beneficio, ha demostrado *la* función, o al menos *una* función del proceso". *Cf.* también Elster (1978a), cap. v.

[84] Merton (1957), pp. 30 *ss.*

métodos menos valiosos de la ciencia social burguesa, y una total falta de interés en los instrumentos que realmente están como mandados a hacer para el análisis marxista. Con esta segunda afirmación me refiero al descuido de la teoría de los juegos por los marxistas; en mi opinión, tal teoría es indispensable para una teoría de explotación o de dominación.[85] Con la primera afirmación me refiero al uso del análisis funcional, a menudo de la índole vulgar, denunciada por Merton. En los escritos marxistas sobre educación, burocracia y, en realidad, sobre casi todos los temas parece estar implícita una idea reguladora de que "toda institución o pauta de conducta en la sociedad capitalista sirve a los intereses del capitalismo y es mantenida porque sirve a estos intereses". Los marxistas parecen haber perdido su sentido de las ironías de la historia, por las cuales las sociedades pueden generar pautas que conducen a su propia destrucción. Para sustentar esta ingenua clase de funcionalismo, los marxistas han inventado un truco especial, que consiste en manipular la perspectiva del tiempo. Digamos, si las acciones del Estado van en contra de los intereses capitalistas a corto plazo, esto tiene la función de salvaguardar los intereses capitalistas a largo plazo; cara, yo gano, cruz, tú pierdes. (*Véase infra*, la sección 8 del capítulo II para ejemplos de tal razonamiento en Marx). Ahora bien, éste no es sólo un procedimiento arbitrario, porque "cualquier argumento puede ser enfocado a cualquier efecto, haciendo equilibrios con la escala del tiempo".[86] También es teóricamente incongruente, porque el análisis funcional no puede

[85] Para primeros pasos en esta dirección, véase Shapley y Shubick (1967) y Lancaster (1973).
[86] Brooke (1972), p. 93.

invocar estrategias indirectas, como ya explicamos en la sección 3. En la medida en que el Estado se mantiene por los efectos de sus acciones sobre la clase capitalista, los efectos negativos a corto plazo debieran hacerle desaparecer (o cambiar) antes de que llegaran a sentirse los efectos positivos a largo plazo. Sólo unos actores intencionales son capaces de dar un paso atrás para después dar dos pasos adelante, de modo que la distinción entre corto plazo y largo plazo conduce lógicamente a una interpretación conspiratoria de la historia, dada la ausencia de pruebas empíricas de tales atenciones.

Desde luego, también hay muchos casos en que los marxistas han cometido la falacia más sutil de suponer que los efectos que son benéficos (a corto plazo) a la clase capitalista tienden a mantener sus propias causas. O que, dicho de otra manera se pueden *explicar* las pautas por una demostración de que sus efectos son benéficos para esa clase. Tomando solamente un ejemplo, una vez más, al azar, Michael Kalecki llega muy cerca de afirmar que el ciclo de negocios puede explicarse por sus efectos benéficos para la clase capitalista.[87] El pleno empleo durante varios periodos de tiempo es políticamente peligroso, mientras que el desempleo permanente es económicamente peligroso, y de ahí la necesidad de un ciclo de negocios. En realidad Kalecki vacila y oscila entre un enfoque conspiratorio y uno funcional, mientras que la historia, si acaso, justifica la afirmación causal de que el ciclo de negocios puede haber tenido estos efectos benéficos.

[87] Kalecki (1971). Para un libro que abunda en proposiciones de esta clase, *véase* Bowles y Gintis (1976).

II. LA RACIONALIDAD IMPERFECTA: ULISES Y LAS SIRENAS

> ... Pero atadme con fuertes lazos, de pie y arrimado a la parte inferior del mástil para que me esté allí sin moverme... Y en el caso de que os ruegue o mande que me soltéis, atadme con más lazos todavía.
>
> *La Odisea*

1. Introducción

Ulises no era por completo racional, pues un ser racional no habría tenido que apelar a este recurso; tampoco era, sencillamente, el pasivo e irracional vehículo de sus cambiantes caprichos y deseos, pues era capaz de alcanzar por medios indirectos el mismo fin que una persona racional habría podido alcanzar de manera directa. Su situación —ser débil, y saberlo— señala la necesidad de una teoría de la *racionalidad imperfecta* que casi ha sido olvidada por filósofos y científicos sociales. La obra seminal de R. H. Strotz [1] y de George Ainslie [2] ha echado los fundamentos empíricos y conceptuales en que habrá que edificar toda obra posterior. En este ensayo intento dar un primer paso hacia una síntesis. También me esfuerzo por ensanchar la propia base empírica y conceptual, añadiendo algunos ejemplos de campos no considerados en su obra. Strotz trata, principalmente, ejemplos tomados de la teoría de

[1] Strotz (1955-1956), criticado en Pollak (1968) y Shefrin y Thaler (1977).
[2] Ainslie (1975, 1977).

la conducta del consumidor, mientras que Ainslie se basa principalmente en descubrimientos de la psicología experimental. Su obra es resumida y analizada en las secciones 5 y 7, respectivamente, del presente capítulo; también la obra de Ainslie se analiza en las secciones 2 y 9. En la sección 6 el lector encontrará un estudio de las preferencias endógenamente cambiantes, problema íntimamente relacionado con las cuestiones de consistencia dinámica planteadas por Strotz. En las secciones 3 y 4, las críticas de la racionalidad ofrecidas por Pascal y Descartes son presentadas y contrastadas. En la sección 8 el análisis se extiende de la conducta individual a la conducta de grupo; las sociedades, así como los individuos, han encontrado útil atarse ellas mismas por medio de constituciones. El problema de atarse a sí mismo es pertinente a muchos problemas de la filosofía de la mente (¿quién ata a quién?). Y de psicología moral (¿quién tiene el derecho de atar a quién?). Estas preguntas corren por todo nuestro ensayo.

La tesis general que estamos defendiendo es que atarse a sí mismo es un modo privilegiado de resolver el problema de la flaqueza de voluntad; la principal técnica para lograr la racionalidad por medios indirectos. Sin embargo, hay otro camino que se puede tomar. Este otro enfoque consiste, aproximadamente, en una redisposición del espacio interno de la persona, sin ningún mecanismo causal establecido en el mundo exterior. Aquí exploramos tres variedades de este enfoque, y las contrastamos con el método escogido por Ulises: la idea de Ainslie de las apuestas privadas indirectas, la idea de Strotz y Pollak de planeación consistente, y la noción de Charles Taulor, de la responsabilidad para sí mismo. Creo

yo que la estrategia de Ulises está íntimamente relacionada con el enfoque aristotélico de la psicología, mientras que las otras nociones tienen estrechos vínculos con la filosofía existencialista. No negaré que puede lograrse cierto grado de control de sí mismo simplemente ayudándose a sí mismo, pero, como argüiré más adelante, se logran resultados más duraderos actuando sobre el medio circundante.

2. Hacia una definición

Comenzaré dando algunos ejemplos de personas que se ataron a sí mismas, o que se comprometieron previamente, como también diré, como variación estilística. Para dejar de fumar, es práctica común fijar alguna maquinaria causal que añada fuerza a nuestra resolución interna: hablar a nuestros amigos de nuestra intención, para provocar sus comentarios sarcásticos si reincidimos; irnos a pasear a la montaña, para no tener cigarrillos a nuestro alcance; atravesar la calle cuando vemos una tabaquería, para no encontrarnos a la vista de cigarrillos; tomar duchas frías para fortalecer nuestra fuerza de voluntad; someternos a hipnosis, para provocar una aversión al tabaco; hacernos creer a nosotros mismos que más cigarrillos significan la muerte cierta dentro de cinco años.

Dejar de fumar es problema relativamente sencillo, por varias razones. En primer lugar, el éxito temporal implica a menudo el éxito permanente, en el sentido de que la abstención durante un periodo prolongado reduce el afán de fumar. En segundo lugar, la adicción a la nicotina es un rasgo bastante aislado del individuo, con pocos efec-

tos de "derrame" sobre otros rasgos del carácter. La *obesidad* —problema central en la bibliografía del autocontrol de la conducta—[3] difiere en ambos aspectos. Ser gordo es un rasgo de carácter (no sólo un rasgo físico), que también participa en muchos otros rasgos; el deseo de cambiar es un deseo de convertirse en una persona diferente. También la tentación de reincidir se reduce mucho más lentamente que en el caso de la nicotina. En un estudio reciente,[4] las técnicas para el autocontrol del peso fueron divididas en dos categorías principales: las técnicas de control de estímulo antecedente, autoaplicadas, y el control de refuerzo, autoiniciado. Dentro de cada una se distinguen otras tres subcategorías, una de las cuales se refiere a la manipulación del medio circundante. Según otro estudio[5] este parece ser, en realidad, el método más eficiente.

Las personas con enfermedad mental periódicamente recurrente pueden atarse a sí mismas de antemano (pero rara vez lo hacen) dando instrucciones de que cuando ocurra el siguiente episodio, las instrucciones que para entonces se habrán dado (negarse a ser hospitalizado u órdenes extravagantes) no sean obedecidas. La Ley de protección de la salud física de Noruega[6] tiene el rasgo exclusivo de que una persona puede buscar, voluntariamente, admisión irreversible en un hospital para enfermos mentales. Para ser precisos, el director de medicina podrá plantear la

[3] *Véase,* Foreyt (comp.) (1977).
[4] Jeffrey (1977).
[5] Stuart (1979).
[6] Debo mi información acerca de esta institución y su unicidad, a Helge Waal y Einar Kringlen, quienes también hicieron valiosas sugerencias sobre otros aspectos. *Véase,* también Adserballe (1977, pp. 485 *ss.)* para una práctica parcialmente similar en Dinamarca.

condición, para su admisión, de que al paciente no le sea permitido salir en tres semanas a partir de la fecha de admisión, aun si desea hacerlo.

La espontaneidad de los músicos de jazz es un fenómeno frecuentemente citado, y que en algunos casos cuenta con buenos testigos. Algunos toman precauciones contra su tendencia a gastar inmediatamente todo dinero que cae en sus manos; otros no lo hacen. En la primera categoría se encontraba Johnny Hodges, quien durante toda su carrera con la banda de Ellington insistió en que se le pagara día por día.[7] En la última categoría estuvo Charlie Parker: "Hombre que vivía de un momento a otro. Hombre que vivía por el principio del placer, la música, el alimento, el sexo, la droga, las sensaciones, su personalidad detenida a un nivel infantil."[8] Esta diferencia entre los dos saxofonistas es fácilmente detectable en su modo de tocar.

Los lectores de *Lucien Leuwen* recordarán el momento en que Madame de Chasteller reconoce su amor a Lucien y, temiendo lo que pueda moverle a hacer, se ata a sí mismo, escogiendo un compañero que ciertamente no le permitirá la menor indiscreción "Cet être si méchant me répondra de moi-même."[9] Podemos hacer aquí una pausa para observar las sutilezas que casi siempre surgen en estos casos. Madame de Chasteller también aduce otro motivo para elegir un compañero: el deseo de castigarse a sí misma por sus pasadas indiscreciones. "Et la sévérité de cette punition tranquillisa sa conscience: madame de Chasteller se pardonna presque l'entrevue

[7] Jewell (1977), pp. 140-141.

[8] Russell (1973); p. 232; también hay abundantes pruebas en Reisner (1974).

[9] Stendhal (1952), I, p. 969.

si légèrement accordée à Leuwen." [10] El castigo de sí mismo puede ser en realidad una técnica de control de sí mismo,[11] pero en este caso sería redundante, porque el castigo también tiene el efecto de hacer físicamente imposible la conducta en cuestión. Si Madame de Chasteller se hubiese castigado por otros medios, como darse azotes, podríamos atribuirle el prudencial motivo de despertar un proceso de aprendizaje operante, aunque difícil habría sido no sospechar que intervenía también algún componente neurótico. En cambio, con la "estrategia del compañero", la prudencia ha conducido a la técnica más sencilla de reducir el número de opciones disponibles, y luego, el motivo para el castigo sólo puede ser neurótico o moral.

Dados estos primeros ejemplos —daremos muchos más— podemos pasar a una definición tentativa de lo que es atarse a sí mismo. Las normas planteadas más adelante bien pueden resultar innecesarias o insuficientes para un planteamiento preciso de nuestras nociones intuitivas del compromiso previo, pero tendrán que bastar como primera aproximación. Un obvio requerimiento es que

i) Atarse a sí mismo es llevar a cabo cierta decisión en el tiempo t_1 para aumentar la probabilidad de llevar a cabo otra decisión en el tiempo t_2.[12]

[10] *Ibid.*

[11] *Véase,* por ejemplo, Mahoney (1972); y también la introducción de los editores a Mahoney y Thoresen (comps.) (1975).

[12] "... el control de sí mismo es sencillamente la emisión de un conjunto de respuestas destinadas a alterar la probabilidad de ocurrencia de otro conjunto de respuestas" (Stuart, 1971, p. 180).

Aquí, el punto fundamental es que el cambio esperado de la probabilidad de la acción ulterior deberá ser el *motivo* de la primera; no un efecto inesperado, ni un efecto predecible, pero recibido sin desagrado. La necesidad de excluir este último caso puede verse considerando el siguiente experimento mental. Si Madame de Chasteller hubiese escogido su compañero para castigarse a sí misma; si, además, el efecto atador de tal elección fuese predecible y aceptable para ella; y si, finalmente, ella no hubiese escogido a esta desagradable persona de no haber cometido sus anteriores indiscreciones, aun suponiendo que pudiera predecir las posteriores, entonces esta primera norma no quedaría satisfecha y no nos encontraríamos ante un caso de atarse a sí mismo. Si la actitud psicológica que acabamos de describir parece inverosímil, yo sólo podré responder que una actitud semejante parece subyacer a la negativa de muchas personas a dedicarse a una conducta complicada y estratégica dirigida contra ellas mismas. Es decir, para algunas personas, el deseo de dejar de fumar puede ser más débil que su deseo de no ser (o llegar a ser) el tipo de persona que lograría el estado de no fumadora atándose a sí misma. Semejante persona se dedicaría a alguna actividad, sabiendo bien que conduciría al estado de no fumadora, que desea; pero la condición para tal actividad puede ser la existencia de algún otro fin que es facilitado por ello y que, en sí mismo, sería motivo suficiente para buscarlo.

Podemos observar que éste *no* es el problema de que algunos resultados (deseados) puedan ser imposibles de alcanzar, salvo como subproductos de actividades emprendidas con otros fines. La persona que se propone obtener placer o conver-

tirse en persona culta habitualmente se verá frustrada, a menos que, en cierto punto, los propios medios se conviertan en fines. *(Cf.* también la sección 9 para otros comentarios sobre "desear lo que no puede desearse".) Tampoco es cuestión de un análisis de costos-beneficios en el estrecho sentido de que los cálculos y deliberaciones necesarios requerirán más tiempo y energía de los que se ahorraría en caso de triunfar la estrategia. Antes bien, la cuestión es de cálculo de costos-beneficios en un sentido mucho más lato: aun si las deliberaciones logran modificar la conducta o el carácter de la manera deseada, *la actividad misma de deliberar* puede modificar el carácter, para empeorar, y en formas consideradas aún más importantes, mediante un efecto funesto sobre la espontaneidad. Es posible volverse adicto a la fuerza de voluntad, y puede ser tan racional tomar medidas contra esta adicción como de cualquier otra. Como lo observa Thomas Nagel en un contexto un tanto diferente, "espontaneidad e inmediatez tienen un valor en sí mismas" [13] y por ello requieren una protección prudencial. Su argumento es que un exceso de espontaneidad puede reducir, hoy, las posibilidades de una conducta espontánea, más adelante; mi argumento es, antes bien, que una estrategia para reducir las consecuencias indeseables de la espontaneidad también puede reducir la cantidad general de espontaneidad en mi conducta.

La condición *(i)* permite también métodos complejos y jerárquicos, que incluyen tres o más decisiones. Por ejemplo, si la pauta de episodios de depresión maniaca, fuese de depresión sucedida por exaltación (en realidad, la pauta usual es la opuesta), entonces podemos suponer el caso

[13] Nagel (1970), p. 73.

siguiente. En el tiempo t_1 la persona, encontrándose en el resultado "normal" o "básico", da la orden de que no se obedezcan las instrucciones que emitirá en t_2, durante la fase depresiva, con el fin de que las que dará en t_3, durante la fase maniaca, no sean obedecidas. Esto presupone que su opinión bien considerada es que una persona en la fase maniaca no necesariamente es irresponsable, o que tratarse a sí mismo como irresponsable agravará las dificultades a largo plazo, aun si las alivia a corto plazo. En términos un tanto fantásticos podemos hablar aquí de una alianza entre el yo temprano y el yo posterior, contra el yo intermedio y más dócil. (Las técnicas de tres pasos también son exploradas por Ainslie, pero en sus ejemplos, el argumento es, antes bien, que la etapa primera y la ulterior pueden combinarse con la intermedia, para hacer superfluo todo compromiso previo.)[14] En este caso, el "estado básico" puede distinguirse claramente de los patológicos, de modo que los observadores externos sabrán qué instrucciones seguir, pero esto puede no ser siempre así. Si una persona emite un día una instrucción en el sentido de que no

[14] Ainslie (1975, p. 478) ofrece, entre muchos otros el siguiente ejemplo: "Otro ejemplo podría ser otra persona con impulsos antisociales, que ha descubierto que los actos de delincuencia lo colocan en situaciones ingratas y por ello trata de evitar la tentación de efectuarlos. Sin embargo, si tiene miedo de perder el control y meterse en mayores dificultades, puede dejar de evitar la tentación de actuar en formas más pequeñas, de modo que las autoridades (policía, ayudantes de pabellón, etcétera) harán mayor esfuerzo por guardarlo. El deseo a largo plazo de evitar un mayor daño le hace buscar recursos que constriñan su futura conducta; dado que tal recurso debe elegirse pronto, cuando la eficacia de la recompensa a la que conduce es baja, también puede ser elegible una que produzca una emoción inmediata, lo que acaso no sea una que se sostenga por sí sola."

se obedezcan sus futuras instrucciones de cierta índole específica, y luego retracta las metainstrucciones al día siguiente, ¿cuál será la norma para decidir si el ego que hace la retractación es más auténtico que el que emitió las metainstrucciones, o el mismo ego cuya falta de autenticidad hizo necesarias las metainstrucciones? En la sección 9 volveré a algunos aspectos de este problema.

La siguiente norma es más ad hoc, pero parece necesaria por la siguiente razón. En la primera norma, tomada en sí misma o en unión con las normas *(iii)*, *(iv)* y *(v) infra*, todo acto de inversión —es decir, todo sacrificio de bienes actuales para disponer después de más bienes— contaría como atarse a sí mismo. Esto, creo yo, es contraintuitivo. Puede ser necesario atarse a sí mismo para hacer ese sacrificio (*véase* el análisis de la norma *(iv)*), es decir, tomar una decisión en t_1 para aumentar la probabilidad del sacrificio en t_2, pero el sacrificio en t_2 no necesariamente es un acto de atarse a sí mismo aun cuando aumenta la probabilidad de llevar a efecto ciertas decisiones de consumo, en t_3. Para excluir estos casos, y otros similares,[15] impondré el requerimiento de que

ii) Si el acto del primer tiempo tiene el efecto de provocar un cambio en el conjunto de opciones de que se dispondrá en momento posterior, entonces esto no cuenta como atarse a sí mismo si el nuevo grupo factible incluye al anterior.

En la sección 9 se arguye que producir cambios en el conjunto factible es una de las diversas es-

[15] Por ejemplo, la idea de Kotarbinski de *planeación flexible*, expuesta en Johansen (1977), pp. 119 *ss*.

trategias de que disponemos para atarnos a nosotros mismos; el segundo requerimiento dice que las expansiones del conjunto factible no contarán como ejemplos de *esta* estrategia. Hay que *excluir* algunas opciones, si queremos hablar de compromisos previos. Una vez más, debemos insistir en la distinción entre motivos y efectos deseados predecibles. Muchos actos de inversión son irreversibles, y como tales tienen la consecuencia de atar al inversionista a alguna asignación particular de consumo, con el tiempo (sección 5, *infra*). Esto hasta podría ser parte del motivo para la inversión, pero no necesariamente. Para ver que una inversión no es irreversible de modo necesario basta considerar los comunes casos "austriacos" en que se deja vino a madurar, o bosques, para que crezcan. Tales actos de "sacrificios" o de "espera" sí producen una expansión en el conjunto futuro de opciones, pero no implican ningún compromiso más allá del actual, porque se puede beber vino y se pueden cortar árboles antes de que lleguen a la madurez.

Una nueva norma es que

> *iii)* El efecto de llevar a cabo la decisión en t_1 debe consistir en desencadenar algún proceso causal en el mundo exterior.

Esto excluye, por ejemplo, las decisiones de decidir. Aquí, la idea no es que cumplir con una decisión de decidir es llevar adelante sencillamente la propia decisión ulterior. Yo considero que una decisión de decidir puede llevarse a cabo estableciendo cierta actitud mental que podemos describir como "resolución", "firmeza de propósitos", etcétera. Y tampoco la razón en que se apoya la norma *(iii)* que las decisones de decidir

carecen de eficacia, aunque yo creo que en general, ejercen muy poco efecto. "Yo decido que yo decidiré que *p*" tiene el mismo sonido ritual y redundante que "si alguien comprara varios ejemplares del periódico matutino para asegurarse de que lo que dice es verdad".[16] El brillante análisis que hace Sartre del jugador, cuya decisión de dejar de jugar pierde toda eficacia causal cuando se encuentra ante "el tapete verde" conduce a la misma conclusión.[17] La necesidad de la tercera norma es, antes bien, que nuestra noción intuitiva de lo que es atarse a sí mismo, a menudo requiere que depositemos temporalmente nuestra voluntad en alguna estructura externa; que establezcamos un proceso causal en el mundo externo que después de algún tiempo retorna a su fuente y modifica nuestro comportamiento. Una vez más, debemos recordar la primera norma, *supra*. La idea no es sencillamente que "un ser humano es controlado por el medio, y también él controla una parte del medio que les afecta a él y a otros".[18] Modificar el medio de un modo que, de manera no anticipada ejerza un efecto ante el modificador no es atarse a sí mismo. Estamos hablando aquí de controlar al controlador *para* controlarse a sí mismo.

La norma *(iii)* también excluye lo que Ainslie llama "apuestas privadas indirectas", la decisión de agrupar las futuras recompensas, de modo que se sostengan o desaparezcan juntas.[19] Esta re-

[16] Wittgenstein (1953), § 265.
[17] Sartre (1943), pp. 69-70.
[18] Jeffrey (1975), p. 194.
[19] Ainslie (1975), pp. 478 *ss.*; Ainslie (1977), pp. 18 *ss.* La base técnica de esta noción es la siguiente. Suponemos que la actual eficacia de la recompensa futura es una función decreciente de tiempo desde ahora hasta el momento de la recompensa. Suponemos, además, que

disposición contable puede tener, de hecho, cierta importancia, según lo he aprendido por experiencia personal. Las probabilidades de que efectuemos ahora alguna tarea desagradable en lugar de aplazarla para el futuro aumentan si nos decimos a nosotros mismos que nuestras razones para aplazarla serán igualmente fuertes cuando llegue la hora futura. Un fracaso actual predecirá fracasos futuros, de modo que las consecuencias de la elección se vuelven más graves que si consideramos sólo la secuela inmediata. Sin embargo, no creo que a esta ingeniosa estrategia se la deba llamar compromiso previo. El hecho de que es una respuesta al mismo problema —flaqueza de voluntad— que motiva la técnica de atarse a sí mismo, no debe movernos a decir que es un ejemplo de tal técnica. Las apuestas privadas indirectas, como la "estrategia de la planeación congruente" que consideraremos en la sección 5 son una alternativa al método de compromiso previo.

la curva que relaciona el tiempo con la eficacia es más cóncava que una curva exponencial (para la naturaleza privilegiada de las curvas exponenciales, *véase* sección 5, *infra*). Esto significa que la eficacia relativa de una gran recompensa en el futuro distante y una recompensa más pequeña en el futuro cercano pueden cambiar cuando se cruzan las curvas correspondientes, lo que nunca pueden hacer si son exponenciales. Hay un punto de cambio, es decir, cuando la persona deja de preferir la gran recompensa y empieza a preferir la más pequeña. Ainslie imagina entonces un caso en el que al sujeto se le ofrece una elección entre dos de tales recompensas, cierto número de veces en sucesión: la apuesta privada indirecta es entonces una decisión de elegir entre todas las parejas simultáneamente en lugar de hacer una sucesión de elecciones. Puede mostrarse entonces que este procedimiento de decisión retrasa el cambio y aumenta la probabilidad de que se elija la recompensa más grande. *Véase* también, la n. 42, *infra*, para la idea de una apuesta *pública* indirecta.

Un cuarto requerimiento, no menos importante, es que

iv) La resistencia contra llevar a efecto la decisión en t_1 debe ser menor que la resistencia que se habría opuesto a llevar adelante la decisión en t_2, si no hubiese intervenido la decisión en t_1.

Darme duchas frías para desarrollar la fuerza de voluntad que me permita dejar de fumar no es muy buena estrategia si colocarme bajo la ducha fría requiere la misma fuerza de voluntad que, supuestamente, desarrollará. El problema que subyace a la condición *(iv)* es la dificultad misma a la cual es solución la estrategia de Ulises: nuestra resistencia general a caminar cuesta arriba, y nuestra preferencia por los paseos cuesta abajo. En la sección 3 del capítulo I, he argüido que la capacidad de *aguardar* y de utilizar *estrategias indirectas* ("un paso atrás, dos pasos adelante") se encuentra entre los rasgos que distinguen al hombre de otros animales, pero esto no es decir que la capacidad siempre es plenamente utilizada. Sólo el hombre está en posesión de la capacidad generalizada de buscar máximos globales a expensas de máximos locales, pero es igualmente cierto que sólo el hombre es susceptible de acracia, porque esta noción sólo tiene sentido ante un trasfondo de triunfante planeamiento a corto plazo y triunfal resistencia a la tentación en muchos casos. *A fortiori,* sólo el hombre es capaz de superar la flaqueza de su voluntad mediante el uso de técnicas como compromiso previo, apuestas privadas indirectas, etcétera.

Esto implica que el compromiso previo en algunos casos puede considerarse como estrategia

indirecta que nos permite utilizar estrategias indirectas. Si soy incapaz de dar todo un paso atrás, entonces tal vez seré capaz de dar medio paso atrás, sabiendo que después del primer medio paso me veré comprometido con el segundo medio, y capacitado a dar, después, dos pasos adelante. Como hemos observado antes, la inversión constituye un ejemplo. Si he diseñado un nuevo tipo de red de pescar que me permitirá atrapar doble cantidad de peces que la red antigua, y si yo sé que mientras hago la nueva red pasaré tanta hambre que preferiré ir a pescar con la red antigua y, por consiguiente, nunca tendré tiempo suficiente para trabajar ininterrumpidamente en la nueva, entonces la *destrucción de la vieja red* puede ser una elección racional: una estrategia indirecta que me obligue a emplear la estrategia indirecta de invertir en una nueva red. Esto *no* es un caso de quemar nuestros puentes, pues si alguien intenta avanzar éstos puentes no servirán, de todos modos, a ningún propósito útil. Destruir medios útiles de producción que habrían podido facilitar las cosas durante la construcción de los nuevos es un procedimiento más radical. Si la necesidad es la madre de la invención, entonces podemos espolearnos a la actividad poniéndonos en situación menesterosa. A veces se ha sugerido que el rápido desarrollo económico de sociedades que fueron aniquiladas por la guerra, como Japón y Alemania después de 1945, debe considerarse bajo esta luz. Yo soy más que escéptico hacia esta idea. Puede ser verdad que la destrucción de maquinaria anticuada tenga el efecto de permitir a una nación escapar de los castigos de ponerse a la delantera, pero de todas las interpretaciones de la historia, salvo la conspiración, ésta nunca debe ser un *motivo* para em-

prender una guerra, y tendría que serlo para satisfacer la primera norma, *supra.* (En la sección 5 del capítulo I, trato [por implicación] de la sugestión aún más oscura de que tal compromiso previo puede ser una *función* de la guerra; volveré a un problema similar en la sección 8.)

El problema general es el siguiente: para una tarea determinada de caminar cuesta arriba (es decir, una tarea a la que yo me resisto con cierta fuerza) y para una cierta fuerza de voluntad, ¿existe siempre una estrategia de compromiso previo que *i)* esté al alcance de mi fuerza de voluntad y que *ii)* sea capaz de llevarme, a la postre, a la cumbre? ("A la postre" porque podríamos pensar en una jerarquía secuencial de actos de atarse a uno mismo, cada uno de los cuales requiera un esfuerzo relativamente pequeño.) Descartes (sección 4, *infra)* arguyó que cualquiera puede lograr cualquier cosa si lo hace de una manera lo bastante indirecta. En un caso especial y en un sentido especial (sección 6, *infra)* esto sigue siendo demostrable aun para personas con ninguna fuerza de voluntad. Si aceptamos, al menos en gracia al argumento, que en este sentido somos todopoderosos, entonces, ¿cómo explicaremos el hecho de que no siempre empleamos las opciones que disponemos? Por hipótesis, no podemos invocar flaqueza de voluntad. Sin embargo, sí podemos invocar consideraciones de costos-beneficios, sea en el sentido lato o en el sentido estrecho que antes distinguimos. Para algunas personas, el acto mismo de deliberar puede requerir un esfuerzo que se encuentra más allá del umbral de su fuerza de voluntad. *(Cf.* el demonio de Maxwell, que sólo podía sobreponerse a la segunda ley de la termodinámica sobre la suposición de que el cálculo no tiene costos ni

fricciones.) Para otras personas, el proceso de ascender (lo que desean) también produciría algún cambio de carácter (a lo que se resisten), caso en el cual podría ser racional rechazar la opción que, supuestamente, existe.

La última condición que impondré es

v) El acto de atarse a sí mismo debe ser un acto de comisión, no de omisión.

La distinción entre omisión y comisión es, desde luego, notoriamente difícil de establecer con precisión, pero, para el caso, no estoy esforzándome mucho por lograr precisión. Observemos primero que cuando alguien no decide hacer *x*, esto, en sí mismo, no cuenta como un acto de omisión; debemos añadir que también decide no hacer *x*. En segundo lugar, consideremos un ejemplo[20] que muestra que podemos atarnos por omisión: un hombre en una esquina de la calle, desesperado por fumarse un cigarrillo ve cómo el encargado de la tabaquería cierra la tienda, se bloquea todo acceso a los cigarrillos, a no entrar. Tales casos quedan excluidos por la norma *(v)*, pero sólo como subproducto no intencional de otros casos que, en mi opinión, definitivamente deben quedar excluidos. Esto muestra que la norma *(v)* puede y debe ser refinada, pero como ya lo dije, no pretendo aquí dar condiciones suficientes y necesarias sino sólo bosquejar los lineamientos generales de la idea.

Para un caso que muestra la necesidad de algo como la norma *(v)*, consideremos el siguiente problema de la filosofía de la educación. Yo sospecho que una visión no insólita del desarrollo de las facultades morales e intelectuales del niño es

[20] Me lo sugirió Robert Goodin.

algo parecido a una secuencia de cuatro etapas. En una primera etapa, el poder de los padres y del maestro no es más que un hecho bruto, aceptado tan sólo porque no hay alternativa. En una segunda etapa, el niño sigue siendo impotente de rechazar la autoridad, pero ya es lo bastante racional para ver que no habría deseado hacerlo aun si hubiese podido. En una tercera etapa, el niño es física y jurídicamente capaz de rechazar la autoridad, pero se retiene de hacerlo porque ve que va en su propio interés el permanecer "atado". En una última etapa, la liberación ocurre, a iniciativa común de todas las partes. Ahora bien, cualquiera que sea el atractivo moral de este cuento de hadas no deseo considerar a la tercera etapa (y, *a fortiori*, no la segunda) como ejemplo de atarse a sí mismo. El hecho de que alguien prefiera no abandonar un estado determinado, no es prueba de que libremente habría entrado a ese estado, a partir de todos los demás estados que estaban abiertos a él. Hay costos de transacción e incertidumbres que destruyen la aparente simetría de las entradas y salidas. Como analizaremos en la sección 6, las preferencias siempre son relativas a una pasada historia de elecciones, y si el niño hubiese sabido por experiencia los estados a los cuales prefiere el estado de ser atado, sus preferencias habrían podido ser muy diferentes. No me parece adecuado decir, como Gerald Dworkin, que "una importante limitación moral al ejercicio de tal poder parental... nos lo ofrece la noción de que el niño llegue, con el tiempo, a ver lo correcto de la intervención de sus padres",[21] pues virtualmente cualquier es-

[21] Dworkin (1972), quien emplea el episodio de Ulises como caso paradigmático de paternalismo justificado.

quema educativo podría extenderse de tal manera que produjere su propia justificación *ex post facto*. Antes bien, diré que la marca de una buena educación es que el niño llegue a ver que tal justificación no es posible, pero que los padres de todos modos tenían que hacer *alguna* elección (injustificada).

3. Pascal

Una manera de hacernos actuar de cierto modo consiste en producir una creencia a partir de la cual la acción se derivará, obligatoriamente. En esta sección, analizaré la lógica de tales *decisiones de creer*, con particular referencia al argumento de la apuesta de Pascal. Este argumento tiene dos partes. La primera se desarrolla, aproximadamente, así:[22] puesto que hay una cierta probabilidad positiva de que Dios exista y dado que quien cree en la existencia de Dios recibe una ganancia infinita si resulta que tiene razón, mientras que sólo tiene en juego una cantidad finita, el principio de la maximización de la utilidad esperada requiere que creamos. (Esto, desde luego, presupone que la dicha eterna no queda reducida al valor actual a alguna positiva tasa de descuento, como en una extravagancia reciente.)[23] Ahora bien, "creencia", es un tipo muy particular de acción, ya que no se le puede efectuar sencillamente por la voluntad de decirlo así. De ahí la necesidad del segundo paso en el argumento:

[22] Para una exposición lúcida, *véase* Hacking (1975), cap. 8.

[23] Azzi y Ehrenberg (1975). Para otro argumento que podría invocarse para mostrar que la dicha eterna equivale a un total finito, *véase*, Elster (1975a), p. 148.

> Queréis llegar a la fe y no conocéis el camino; queréis curaros de la infidelidad y solicitáis el remedio: aprended de quienes han estado atados como vosotros y que ahora ponen en juego todo lo que tienen; son gente que conoce este camino que quisiérais seguir, y que están curadas de un mal que queréis curaros. Seguid la manera como han comenzado; haciéndolo todo como si creyeran, tomando agua bendita, haciendo decir misas, etcétera. Naturalmente, hasta esto os hará creer y os embrutecerá. —Pero esto es lo que temo. ¿Y por qué? ¿Qué vais a perder? [24]

Inicialmente, no hay razón para creer, sólo una razón para hacernos creer. La eficacia causal para una creencia para un fin dado nunca puede ofrecer motivos o razones para *adoptar* esa creencia, con la posible excepción de las creencias que se realizan por sí mismas.[25] Sin embargo, esta efi-

[24] Vous voulez aller à la foi, et vous n'en savez pas le chemin; vous voulez vous guérir de l'infidélité, et vous en demandez le remède: apprenez de ceux qui ont été liés comme vous, et qui parient maintenant tour leur bien; ce sont gens qui savent ce chemin que vous voudriez suivre, et guéris d'un mal dont vouz voulez guérir. Suivez la manière par où ils ont commencé: c'est en faisant tout comme s'ils croyaient, en prenant de l'eau bénite, en faisant dire des messes, etc. Naturellement même cela vous fera croire et vous abêtira. —Mais c'est ce que je crains. —Et pourquoi? Qu'avez-vous à perdre? *Pensée* 233.

[25] Cap. II, sección 9 y en particular n. 123, *infra*, para un breve análisis de este problema. La idea es que en algunos casos la creencia no sólo puede cambiar el mundo de alguna manera (deseada) sino cambiarlo de una manera que hace que la creencia se vuelva verdad. Más formalmente, supongamos que el estado del mundo y es una función f de mis creencias acerca del mundo x (y de un número de otras variables que se mantienen constantes con los propósitos del análisis). Supongamos además que mi utilidad u es una función g de y. También supongamos que las condiciones del teorema del punto

cacia puede constituir una razón para comprometerse previamente con la creencia, en el sentido de fijar una serie de acciones que tendrán el predecible resultado de que lleguemos a creer. En el argumento de Pascal pueden describirse estas acciones diciendo que "pasan por los movimientos"; actuar como si creyéramos, para generar la cosa auténtica.

El argumento de Pascal, dirigido a sí mismo, puede contrastarse con el argumento dirigido hacia otro, que se le ocurre con más naturalidad al científico social. Arthur Stinchcombe nos ofrece el siguiente agudo análisis:

> Hay dos posibles nexos causales entre el control sobre las actividades en el presente y la estructura de las actividades y los autores en el futuro *a)* Las actividades establecidas por actuales detentadores del poder, que encarnan sus valores, pueden servir a otras soluciones que las de servir a tales valores. Estas funciones adicionales conservarán las actividades, aun si el compromiso con el valor es bajo. *b)* Las personas quedan comprometidas con lo que están haciendo, tal vez para reducir la disonancia cognoscitiva, de tal manera que un modo de socializar a la gente es hacerle actuar en términos de tal valor sin fe y permitir que luego la fe siga.[26]

fijo de Brouwer quedan satisfechas *(véase,* Quirk y Saposnik (1968) para una explicación no técnica), de modo que existe una x tal que $x = f(x)$. Definimos entonces x como el valor de x que maximiza $u = g(f(x))$. Adoptar una creencia que se autorrealizara por su eficacia causal requeriría entonces $x = x$, lo que sólo podría ocurrir por accidente. Una segunda posibilidad, que también podría ocurrir por accidente, es que hay varios puntos fijos, de los cuales escogeríamos el que maximizara $g(f(x))$. Esta posibilidad vuelve a plantearse en la sección 9 del capítulo II, *infra.*

[26] Stinchcombe (1968), p. 116.

El segundo de estos dos nexos causales también se encuentra en el meollo del argumento de Pascal, con la diferencia —enorme— de que en su caso, a los *propios* individuos se les está pidiendo que manipulen sus propias creencias, con la acción como eslabón intermedio. Desde luego, es bien sabido que los gobernantes o los clérigos pueden explotar la pompa de las ceremonias religiosas para robustecer la fe religiosa contra el ataque de la razón; tomemos un característico pasaje de Stendhal: "Un *Te Deum*, nubes de incienso, infinitas descargas de mosquetería y de artillería; los campesinos estaban embriagados de dicha y de piedad. Una jornada semejante deshace la obra de cien números de periódicos jacobinos." [27] Mucho más paradójico es que la razón deba adoptar tales métodos contra sí misma. Valéry, en su maravilloso ensayo sobre Stendhal, entró sobre una mala pista cuando arguyó que Stendhal habría considerado odiosa la (segunda parte de) el argumento de la apuesta.[28] Cierto es que Stendhal detestaba la imitación pasiva y mecánica que es la base de la religión y de la tradición, pero la exhortación a la imitación *activa* se dirige precisamente al libertino inteligente, de las opiniones de Stendhal.

Miremos más de cerca las paradojas de la decisión de creer. Bernard Williams ha dicho, con-

[27] Stendhal (1952), I, pp. 317-318.

[28] "A des hommes de cette espèce, traditions et religions son antipathiques para essence et même odieuses. Ils y voient des puissances fondées sur *l'imitation*, et cette imitation renforcée au besoin, comme le marque et le conseille fort bien Pascal, par la *comédie:* 'Suivez la manière par ou ils ont commencé; c'est *en faisant tout comme s'ils croyaient*, en prenant de l'eau bénite', etc. Imaginer ici le visage de Beyle lisant cette phrase, si jamais il l'a lue" (Valéry, 1957, p. 578).

vincentemente,[29] que aun si es posible decidir creer que *p*, no es posible a la vez creer que *p* y creer que *p* se basa en una decisión de creer que *p*. Si la decisión de creer que *p* se lleva adelante con éxito, también *p* debe borrarse a sí misma de la memoria del creyente. La idea no es que la creencia en que *p* es incompatible con la creencia en que la creencia de que *p* es el resultado de algún proceso causal; puede presumirse que todas las creencias son producidas causalmente. La idea es, antes bien, que en el caso de una decisión de creer, la creencia brota del tipo erróneo de proceso causal, es decir, de un proceso no relacionado con algún motivo o razón para creer.[30] La implicación de este argumento es que la decisión de creer sólo puede llevarse a cabo de modo satisfactorio si va acompañada por una decisión de olvidar, es decir, la decisión de olvidar la decisión de creer. Sin embargo, esto es tan paradójico como la decisión de creer:

> El Corazón no puede olvidar
> a menos que contemple
> lo que declina.[31]

[29] B. A. O. Williams (1973); y Pears (1974, pp. 105 *ss.*), quien señala un problema que cuando se aplica a Pascal, puede formularse así: en el proceso gradual de una creencia creciente y de una razón menguante, ¿no puede surgir un punto en que la primera aún no es lo bastante fuerte para mantener la creencia religiosa, y la segunda ya no tan fuerte que pueda hacerlo? Un caso bastante diferente en que podría parecer racional o deseable manipularse a sí mismo para creer algo para lo cual no se tienen motivos racionales de fe es el llamado Problema de Newcomb, para el cual remitimos al lector al cap. IV, sec. 3, *infra* y a Elster (1978a), cap. IV.

[30] *Cf.* Davidson (1970) para esta noción.

[31] Dickinson (1970), núm. 1560. B. A. O. Williams (1973, pp. 150-151) ve una asimetría entre decidir olvidar y decidir creer, pero su argumento parece basarse en una con-

Podemos *hacer* que alguien olvide algo, pero no diciéndole que lo olvide, lo que muy probablemente tendrá el efecto contrario, si toma la orden en serio.[32] *A fortiori,* tampoco podemos sencillamente decidir olvidar algo; a menos que, asimismo, se ate usted en alguna manera, poniendo en movimiento alguna maquinaria indirecta que produzca el olvido. El procedimiento más eficiente sería el iniciar un proceso causal sencillo con el doble efecto de provocar creencia y de hacer olvidar que alguna vez empezó. Pedir que nos hipnoticen es uno de tales mecanismos. Actuar como si uno creyera es otro, en la medida en que "cela vous fera croire *et* vous abêtira". La pérdida de la facultad crítica no es simplemente un *subproducto* de la fe inducida por sí misma, sino una condición esencial para que tal fe sea tomada en serio, y si no fue seguida a partir del proceso mismo de inducción de la fe, se habría necesitado un proceso separado para causarla.

Ahora bien, aun si esto puede conducir a que la fe sea tomada en serio, muchos se han preguntado si semejante fe debe tomarse en serio.[33] Tendría que ser un dios muy tolerante y con

fusión entre una decisión de olvidar algo que yo he sabido y la decisión de no adquirir cierto conocimiento nuevo sin el cual pienso que puedo pasármela. No hay razón por la cual yo deba siquiera tratar de *completar* el conocimiento, pero hay buenas razones para una adquisición *acumulativa* de conocimiento. Cierto es que hay muchas piezas adquiridas de conocimiento que son tan poco importantes que con toda tranquilidad puede olvidarlas, pero por su posición esto no es verdad de un caso de información que es considerado digno de ser deliberadamente olvidado. Este es el punto del poema de Emily Dickinson.

[32] *Cf.* III.9 y cap. IV, *infra,* para más acerca de estas paradójicas órdenes de la variedad de doble atadura.

[33] James (1896).

nulo sentido crítico para conceder su gracia a quien adquirió su fe de esta manera; parecería mucho a elogiar a alguien por tener la razón mientras se sabía que tenía razón por las razones erróneas. Ahora bien, después de los hechos (es decir, tras la llegada de la fe) podemos decir de alguien para quien la práctica religiosa de actuar como si fuera creyente sólo tuvo el efecto de desencadenar una fe que en retrospectiva estuvo latente todo el tiempo, y podríamos argüir que la génesis de la fe y el hecho de que fue adquirida por razones erróneas no tendría importancia si, en realidad, es sostenida por razones correctas.[34] Sin embargo, este es un argumento con truco. Supongo que nadie podría aceptarlo por sí mismo antes de los hechos, es decir, en el momento en que la apuesta debía ser aceptada o rechazada; pues o bien esta afirmación no sería más que un caso de autoengaño, que no engañaría a Dios, o sería auténtica, que sólo podría ser aceptada por alguien que ya hubiese adquirido una fe auténtica. Tal vez sea posible que alguien se levante tirando de sus cordones de sus botas, pero nadie puede contar, racionalmente, con ser capaz de hacerlo.

A menos que esté yo en un error, hay en Pascal una segunda línea de argumento que es un tanto difícil de embonar con la apuesta. En el argumento de la apuesta se supone que la razón en sí misma no puede ofrecer ningún argumento en favor de la existencia de Dios, sólo en favor de la utilidad de creer en esa existencia. Sin embargo, en otros contextos, Pascal parece creer que la razón es capaz de llegar a esa creencia mediante raciocinio, pero que necesita la ayuda de las pasiones para *mantener* la fe:

[34] Cargile (1966).

Pero hay que desengañarse: tenemos tanto de autómata como de espíritu; y de aquí viene que el instrumento por el cual se produce la persuación no sea únicamente la demostración. ¡Qué pocas cosas demostradas hay! Las pruebas no convencen más que al espíritu. La costumbre hace que nuestras pruebas sean las más fuertes y las más creídas; inclina al autómata que arrastra al espíritu sin pensar en ello... Finalmente, hay que recurrir a la costumbre cuando el espíritu ha visto una vez dónde está la verdad, a fin de abrevar en ella, y asirnos a esta creencia, que nos escapa en todo momento; porque es demasiado trabajo tener siempre presentes sus pruebas. Hay que adquirir una creencia más fácil, la del hábito, que sin violencia, sin arte, sin argumento, nos hace creer en las cosas, inclina todas nuestras potencias hacia esta creencia, de suerte que nuestra alma caiga en ella naturalmente. Cuando no se cree sino por la fuerza de la convicción y el autómata está inclinado a creer lo contrario, no es bastante. Hay que hacer creer, pues, a nuestras dos piezas: al espíritu, por las razones, que basta con haber visto una vez en su vida, y al autómata, por la costumbre, no permitiéndole que se incline hacia lo contrario.[35]

[35] Car il ne faut pas se méconnaître: nous sommes automate autant qu'esprit; et de là vient que l'instrument par lequel la persuasion se fait n'est pas la seule démonstration. Combien y a-t-il peu de choses démontrées! Les preuves ne convainquent que l'esprit. La coutume fait nos preuves les plus fortes et les plus crues; elle incline l'automate, qui entraîne l'esprit sans qu'il y pense... Enfin il faut avoir recours à la coutume quand une fois l'esprit a vu où est la vérité, afin de nous abreuver et nous teindre de cette creánce, qui nous échappe à toute heure; car d'en avoir toujours les preuves présentes, c'est trop d'affaire. Il faut acquérir une creánce plus facile, qui est celle de l'habitude, qui, sans violence, sans art, sans argument, nous fait croire les choses, et incline toutes nos puissances à cette croyance, en sorte que notre âme y tombe naturellement. Quand on ne croit que par la

Esta parece ser, esencialmente, la clásica visión aristotélica de la educación moral: nos volvemos hombres buenos al efectuar buenas acciones. Inicialmente, el desempeñar estas acciones es como trepar cuesta arriba, lo que requiere esfuerzo y tropieza con resistencia, pero el resultado es fijar un *hexis* tal que las mismas acciones fluyan tan naturalmente de él como el agua que corre cuesta abajo.[36] La distinción entre saber algo en abstracto y conocerlo en el sentido de tener la prueba ante nuestro espíritu, también es utilizada por Aristóteles en su solución al problema de la *acracia*. El *akrátes* sabe lo que debe hacer, pero sólo en el sentido en que puede decirse que un geómetra dormido "conoce" un teorema geométrico. No podemos *mantener* constantemente ante nuestro espíritu todo lo que sabemos, aun si logramos *traerlo* a la mente en las condiciones apropiadas, que pueden carecer del calor de la acción. Ahora bien, esta es una "disolución" al problema de la falta de voluntad, tan sólo en el sentido de explicar cómo es posible tal fenómeno,[37] y no una

force de la conviction, et que l'automate est incliné à croire le contraire, ce n'est pas assez. Il faut donc faire croire nos deux pièces: l'esprit, par les raisons, qu'il suffit d'avoir vues une fois en sa vie; et l'automate, par la coutume, et en ne lui permettant pas de s'incliner au contraire. *Pensée* 252.

[36] *Ética nicomaquea* 1103. Una distinción similar se encuentra en Mencio, entre "actuar por medio de benevolencia y justicia" y "poner en acción benevolencia y justicia" (Nivison, 1976). Existen fascinantes analogías y diferencias entre Aristóteles y Mencio (en la interpretación de Nivison) que son dignas de una exploración más detallada y competente de la que puedo ofrecer.

[37] Para análisis de este problema, *véanse* los artículos reunidos en Mortimore (comp.) (1971); también (y esp.) Davidson (1969), quien concluye que la flaqueza de voluntad es una forma de *irracionalidad* en que los procesos causales de la mente operan a espaldas, por decirlo así,

solución en el sentido de una estrategia para superarla. Una solución en este último sentido se presenta si hacemos que la teoría aristotélica de la educación moral pese sobre el problema de la *acracia*; tenemos entonces la opinión expuesta por Pascal en el texto antes citado. Según este parecer, la razón tiene una doble tarea. En primer lugar, sólo la razón puede llegar a la visión intelectual de que Dios existe. En segundo lugar, la razón sabe que la visión intelectual no basta, y por tanto inicia una serie de acciones que engendrarán la creencia "automática" o "habitual", permitiéndome actuar rectamente sin tener que movilizar a cada instante toda la serie de argumentos.

Debe quedar bastante claro que esta línea de razonamiento es muy distinta del argumento de la apuesta. Tratemos de elucidar la diferencia en forma más explícita. Si llamamos creencia_r a la creencia racional a la que se llegó (y se mantuvo) por raciocinio, y creencia_c a la creencia "automática" o habitual y *p* a la proposición "Dios existe", el argumento de la apuesta se desarrolla de la siguiente manera:

(1) Yo no creo_r que *p*.
(2) Yo creo_r que siempre debo creer_c que *p* o creer_r que *p*.
(3) Yo no creo_r que deba llegar a creer_r que *p*.
(4) Yo creo_r que es imposible a la vez creer_c que *p* y creer_r que yo creo_c que *p* por causa de una decisión de creer_c que *p*.
(5) Por tanto decido actuar de tal modo que pro-

del ego deliberante. Las razones que *hacen* que yo haga *x* pueden prevalecer sobre las *razones* que son razones para hacer *y*, aun cuando éstas sean más fuertes (*qua* razones) que las anteriores.

duzca *i)* que yo creo_c que *p* y *ii)* que yo no creo_r que crea_c por causa de una decisión de creer_c que *p*.

Aquí, la premisa (1) es el punto de partida del argumento, que va dirigido al incrédulo. La premisa (2) es la conclusión de lo que yo he llamado la primera parte del argumento de la apuesta, es decir, el razonamiento en términos de la maximización de la utilidad esperada. La premisa (3) se sigue, como antes se dijo, de la premisa (1). La premisa (4) se sigue del argumento de Williams. Las premisas (2) y (3) dan parte *i)* de la conclusión (5); la parte *i)* y la premisa (4) entonces dan la parte *ii)*. Por contraste, el argumento aristotélico dice así:

(6) Creo_r que *p*.
(7) Creo_r que siempre debo creer_c que *p* o creer_r que *p*.
(8) Yo creo_r que es imposible creer_r siempre que *p*.
(9) Por tanto decido actuar en tal forma que siempre crea_c que *p*.

De la fundamental diferencia entre las premisas (1) y (6) surgen varias diferencias más entre los dos argumentos. Las premisas (2) y (7) son idénticas en su redacción, pero la palabra "debo" no refleja el mismo tipo de racionalidad en ambos casos. La premisa (6) es en sí misma una razón suficiente para (7), en tanto que (2) sólo puede sostenerse invocando la eficacia causal de la fe. Las premisas (3) y (4) corresponden en cierto sentido a la premisa (8), pero la lógica es muy diferente. El que (3) se sigue de (1) es conceptualmente necesario, como lo es también la premisa (4), por contraste (8) sólo expresa una

dificultad psicológica. Ambos argumentos concluyen que debemos atarnos a una creencia$_c$, pero el argumento de la apuesta tiene la implicación añadida de que debemos atarnos a olvidar el argumento mismo.

Hay en Pascal una frase que podría usarse como apéndice de ambos argumentos: "No hay nada tan conforme a la razón como ese rechazo de la razón." [38] Ahora, ya es claro que este rechazo puede ser interpretado en el sentido fuerte de una *abdicación de la razón,* o en el sentido débil de la *insuficiencia de la razón,* por sí sola. En la sección 4 veremos que la crítica cartesiana a la razón probablemente es más débil que cualquiera de éstos. Allí insistiremos en el hecho de que hasta la abdicación de la razón es una noción menos radical que el destronamiento de la razón que nosotros asociamos a los nombres de Nietzsche y Freud. En el argumento de la apuesta es la razón misma la que decide abdicar; una racionalidad de segundo nivel, decidiendo que se debe abolir la racionalidad. El derrocamiento de la razón es más radical, porque presupone que todas las formas de la racionalidad no son mas que afeites de los impulsos vitales; en realidad es erróneo hablar incluso del derrocamiento pues la razón nunca gobernó señeramente. El derrocamiento de la razón sólo efectúa *de jure* lo que siempre había ocurrido *de facto:* la supremacía de las pasiones.

4. Descartes

A la frase de Pascal citada al comienzo del último párrafo corresponde esta frase de Descartes:

[38] Il n'y a rien de si conforme à la raison que ce désaveneu de la raison. *Pensé* 272.

"la principale finesse est de ne vouloir point du tout user de finesse".[39] En las siguientes páginas propondré una interpretación de esta frase y, más generalmente, de la teoría cartesiana de la elección racional. Como ya se dijo, la crítica de la razón (*finesse*) implícita en tal afirmación probablemente es bastante débil; se le puede caracterizar en pocas palabras como *Crítica de la racionalidad instantánea.* Yo argüiré que aunque esta crítica no implique la insuficiencia de la razón, no resulta inverosímil ver una conexión psicológica entre las dos. Con este propósito empezaré mostrando que Descartes tiene en realidad una teoría, planteada en *Les passions de l'âme,* que es muy similar al argumento aristotélico aducido por Pascal. Aunque el propio Descartes nunca hizo que esta teoría pesara sobre la teoría de la racionalidad instantánea, el que mantuviera ambas opiniones y el que exista una manera verosímil de relacionarlas parecería justificar la interpretación propuesta hacia el fin del estudio.

A finales de la primera parte de *Les passions de l'âme,* Descartes establece su distinción entre dos maneras de combatir las pasiones:

> ... aquellos en quienes la voluntad puede vencer más fácilmente las pasiones y contener los movimientos de los cuerpos que las acompañan, sin duda tienen las almas más fuertes. Pero hay quienes no pueden sentir su fuerza porque nunca combaten su voluntad con sus propias armas, sino tan sólo con las que les ofrecen algunas pasiones para resistir a otras. Lo que yo llamo sus propias armas son juicios firmes y determinados que tocan al conocimiento del bien y del mal, según los cuales ha resuelto conducir las acciones de su vida.[40]

[39] Descartes (1897-1910), IV, p. 357.
[40] *Ibid.,* xi, pp. 366-367.

La voluntad puede superar las pasiones mediante simple fuerza de voluntad, pero también utilizando la estrategia indirecta de enfrentar las pasiones unas contra otras. Albert Hirschman ha explorado recientemente el difundido uso de esta idea en el siglo XVII,[41] pero tal vez no distingue con claridad suficiente entre los propósitos analíticos, los manipulativos y los estratégicos a los que se la puede unir. El propósito analítico sería determinar en qué grado las pasiones en realidad tienden a neutralizarse unas a otras *en los hombres tal como son.* Los enfoques manipulativos y estratégicos enfrentarían una pasión contra otra para modificar la conducta: la conducta de otros en el caso manipulativo y la conducta de uno mismo en el caso estratégico. Esta modificación podría operar de una de dos maneras. En primer lugar, la persona A puede tratar de influir sobre la persona B (dejando abierta la posibilidad de que A y B puedan ser la misma persona), asociando alguna recompensa o castigo adicional a algunos de los cursos de acción a que tiene acceso B. Para el caso de A = B, este es el método de las apuestas públicas indirectas.[42] En se-

[41] Hirschman (1977), esp. pp. 20 ss.

[42] Una apuesta pública indirecta es un cambio irrevocable en la estructura de recompensas, a diferencia de las apuestas privadas indirectas que pueden invertirse a voluntad. Tal como fue analizado extensamente por T. S. Schelling (1963), las apuestas públicas indirectas pueden utilizarse para *mejorar* mi resultado a partir de una situación similar a un juego, reduciendo el pago que recibo en ciertas combinaciones de estrategia. Desde luego, es decisivo que la apuesta indirecta sea creíble, por lo que mi adversario realmente actúe sobre la suposición de una modificada estructura de pagos. Es un hecho sorprendente y digno de explicación el que ninguna sociedad (hasta donde yo sé) haya desarrollado instituciones, privadas o públicas, especializadas en aceptar apuestas indirectas con

gundo lugar, A puede desear cambiar el carácter (es decir, el sistema de pasiones) de B, de modo que aun con una estructura de recompensa constante pudiera establecerse una elección diferente. Enfrentar tal pasión con tal pasión puede tomar la forma de modificar la situación para que nuevas pasiones influyan sobre ella, o de cambiar las pasiones para que afecten la situación de un modo nuevo. Yo creo que todos los casos aducidos por Hirschman (además de los puramente analíticos) entran en la categoría de manipulación-recompensa. También creo que la idea principal de Descartes es que la modificación de la conducta puede lograrse mediante una modificación de carácter autodirigida (estratégica). El siguiente pasaje debe apoyar esta interpretación:

> Es útil saber que como ya se dijo antes, aunque cada movimiento de la glándula parece unido por la naturaleza a cada uno de nuestros pensamientos desde el comienzo de nuestra vida, sin embargo se les puede unir a otros por hábito... Ahora bien, esas cosas son útiles de saber para dar valor a cualquiera para estudiar y observar sus pasiones; pues, aunque con un poco de industria se puedan cambiar los movimientos del cerebro en los animales desprovistos de razón, es evidente que se le puede hacer mejor aún en los hombres, y que los mismos tienen las almas más cébiles podrían adquirir un imperio muy absoluto sobre todas sus pasiones, si se empleara suficiente industria para amansarlas y conducirlas.[43]

el propósito de acomodar los deseos de la gente para atarse a sí misma. Tales sustituciones podrían ser utilizadas para interacciones estratégicas con otras, o sencillamente, para autocontrol.

[43] Et il est utile de savoir que, comme il a déjà été dit ci-dessus, encore que chaque mouvement de la glande semble avoir été joint par la nature à chacune de nos pen-

Cualquiera puede hacer cualquier cosa: la mínima fuerza de voluntad basta para las hazañas más extraordinarias de dominio de sí mismo, contando con un entendimiento de los mecanismos fisiológicos por los que se forman y cambian los hábitos. *Hexis* en Aristóteles, costumbre en Pascal y hábito en Descartes, son considerados, todos ellos, como resultado final de acciones no habituales. El resultado puede ser no intencional ni imprevisto, si somos ignorantes del proceso casual por el cual se produce, pero si alguien tiene un atisbo del funcionamiento del espíritu (de acuerdo con Aristóteles, debemos tenerlo),[44] puede explotársele con el propósito del dominio propio.

Creo haber mostrado que el tema del compromiso previo está presente, de hecho, en la psicología cartesiana, pero ciertamente es menos importante que el otro tema, que ahora procedo a elucidar. Este tema —la crítica de la racionalidad instantánea— aparece por primera vez en un célebre pasaje del *Discours de la méthode:*

> La segunda máxima de mi moral consistía en emplear en mis actos la mayor energía y firmeza de que fuera capaz y seguir las opiniones dudo-

sées des le commencement de notre vie, on les peut toutefois joindre à d'autres para l'habitude... Or ces choses sont utiles à savoir pour donner le courage à un chacun d'étudier à regarder ses passions; car, puisqu'on peut, avec un peu d'industrie, changer les mouvements du cerveau dans les animaux dépourvus de raison, il est évident qu'on le peut encore mieux dans les hommes, et que ceux même qui ont les plus faibles ames pourraient acquérir un empire très absolu sur toutes leurs passions, si on employait assez d'industrie à les dresser et a les conduire. Descartes (1897-1910), XLI, pp. 368-369.

[44] *Etica Nicomaquea* 1114; *véase* la sec. 6 del presente capítulo para un análisis de este requerimiento.

> sas, una vez aceptadas, con constancia, como si fueran segurísimas, imitando así a los viajeros que, extraviados en un bosque, no deben detenerse ni elegir un camino para luego desandarlo y elegir nuevamente. Deben, por el contrario, escoger el que les parezca conveniente, y seguirlo, sin volver la vista atrás, y si todos les parecen lo mismo, seguir uno cualquiera, pero sin romper un paso, porque si no llegan al sitio que desean, al menos ese camino les conducirá al lugar más seguro que el centro de una selva.[45]

Al interpretar este texto, seguiré un procedimiento en dos pasos. En el primer paso esbozaré cierto número de interpretaciones posibles, dos de las cuales son escogidas como igualmente verosímiles (y como igualmente diferentes de la crítica pascaliana de la razón). En segundo paso, cito algunos pasajes más de Descartes que tienden a atenuar el contraste entre Descartes y Pascal, aun cuando yo creo que queda una auténtica diferencia. A lo largo de la elucidación, me baso en la moderna economía y la teoría de los juegos, que han hecho posible una discusión más precisa de los problemas a los que se enfrentó Descartes.

[45] Ma seconde maxime était d'être le plus ferme et le plus résolu en mes actions que je pourrais, et de ne suivre pas moins constamment les opinions les plus douteuses, lorsque je n'y serais une fois déterminé, que si elles eussent été très assurées. Imitant en ceci les voyageurs qui, se trouvant égarés en quelque foret, ne doivent pas errer en tournoyant tantôt d'un côté, tantôt d'un autre, ni encore moins s'arrêter en une place, mais marcher toujours le plus droit qu'ils peuvent vers un même côte, et ne le changer point pour de faibles raisons, encore que ce n'ait peut-être été au commencement que le hasard seul qui les ait déterminés à le choisir; car, par ce moyen, s'ils ne vont justement où ils désirent, ils arriveront au moins à la fin quelque part où vraisemblablement ils seront mieux que dans le milieu d'une forêt. Descartes (1898-1900), VI, p. 24.

Algunas de las ideas desarrolladas más adelante también se analizan, desde una perspectiva un tanto distinta en el capítulo III, sección 5.

La *primera interpretación* del pasaje que acude a la mente de un lector versado en economía es, creo yo, la siguiente: una continua evaluación y reevaluación requiere de tanto tiempo que pueda esperarse que compense, con creces, el tiempo necesario por la mejorada dirección que viene de la evaluación. El viajero saldrá antes de la selva si sigue una línea recta —arbitrariamente elegida— que si se detiene constantemente para cambiar de dirección. Esta interpretación está muy cerca de los intentos que se han hecho [46] para reducir la teoría de Herbert Simon [47] de la "racionalidad satisfactoria" o "limitada" a una especie de racionalidad maximizadora o total. Esta reducción subraya que el empleo de líneas generales o principios de decisión estereotipados, como "siempre sígase una línea recta cuando se esté perdido en una selva", puede ser óptimo si nos permite economizar el costo y recabar y evaluar información. Por ejemplo, buscar la solución abstractamente óptima de un problema técnico puede exigir tanto tiempo y dinero que la empresa tenga que declararse en quiebra antes de que surja la solución. La razón de que las empresas francesas regularmente pierdan contratos ante sus competidores norteamericanos se ha buscado a lo largo de estos lineamientos; la excesiva racionalidad de los franceses se convierte en un obstáculo para la racionalidad económica. He dicho en otra parte que el racionalismo de Leibniz, o antes bien su intento de trasladar el racionalismo filosófico a la racionalidad económi-

[46] *Véase*, por ejemplo Riker y Ordeshok (1973), cap. 2.
[47] Simon (1954).

ca, también es profundamente irracional desde el punto de vista práctico.[48] Hasta donde sea correcta esta primera interpretación de Descartes, se podría decir que constituye una respuesta anticipada a Leibniz.

Yo creo, en realidad, que esta interpretación es una de las dos que con casi la misma verosimilitud pueden imputarse a Descartes. Antes de pasar a las demás interpretaciones, deseo insistir en que la decisión de seguir una línea recta es una auténtica *decisión*, un acto de comisión, no de omisión. Se hace ante los hechos, y no es una simple racionalización formada (o imputada al actor) tras los hechos, y tiene la ocurrencia del hecho como su prueba única o principal. Este punto es importante en vista de las numerosas pseudoexplicaciones que en años recientes se han ofrecido para probar que todas las conductas aparentemente irracionales "en realidad" están gobernadas por el principio de la elección racional. Veremos este problema general en el capítulo III sección 10. Aquí sólo mencionaré el argumento de Douglas North en el sentido de que una ideología es "un modo de economizar los costos de información y por consiguiente [es] en general una respuesta racional",[49] porque nos ahorra la molestia de evaluar cada situación por separado y sobre sus propios méritos. Ahora bien, concebiblemente, alguien podría adoptar una ideología, como una actitud rígidamente negativa hacia un grupo minoritario por semejantes razones pero, si se quiere que tenga alguna fuerza este tipo de explicación, habríamos de ubicar la decisión en el espacio y en el tiempo. Decir que el hecho de *no abandonar* una ideología que se ha reci-

[48] Elster (1975a), cap. III.
[49] North (1971), p. 122.

bido de nuestros propios padres es prueba de que hemos *adoptado deliberadamente* la ideología para economizar los costos de información no sería una falacia similar a la que analizamos al final de la sección 2 del presente capítulo. (Del análisis de la sección 5 del capítulo I también se sigue que debemos estar en guardia de que la idea de una ideología podría ser racional aunque no fuera deliberadamente adoptada, en el sentido de ser una respuesta *funcional* a un problema dado. La selección natural de ideologías parece una noción sumamente traída por los cabellos, en favor de la cual no hubo ninguna prueba empírica.)

En los dos últimos párrafos hemos distinguido entre cuatro enfoques de la elección racional: la racionalidad maximizadora (Leibniz); la racionalidad satisfaciente (Simon): "la satisfacción como maximización" (primera interpretación de Descartes); "la maximización como satisfacción" (las teorías de la selección natural sobre la conducta racional). En un artículo convincente y en extremo olvidado, Sidney Winter ha ofrecido algunos argumentos importantes que tienden a demoler las tres alternativas al enfoque de Simon; estos argumentos también sugieren la *segunda interpretación* de Descartes. Winter observa que el intento de reducir la satisfacción a la maximización produce una regresión infinita, porque "esta elección de una estructura de información maximizadora de la ganancia requiere información, y no es claro cómo el aspirante a maximizador de la ganancia adquiere esta información, o qué garantiza que no pagará un precio excesivo por ellas".[50] Tomemos el caso de una empresa multinacional que decide no entrar en el merca-

[50] Winter (1964), p. 252; *véase también* Winter (1975).

do de cambios porque los costos de información de la operación excederían los beneficios.[51] Entonces tendremos que preguntar cómo decidió la empresa cuánta información debía obtener antes de llegar a la decisión de no adquirir la información necesaria para las operaciones de cambio. A menos que podamos probar (y no veo cómo lo podríamos hacer) que la desviación del óptimo "real" converge en cero o, en todo caso, pronto se vuelve menor para cada nuevo nivel en la jerarquía de estructuras de información, este argumento no sólo tiene la implicación de que en cada decisión debe haber un punto de corte, donde el cálculo cesa y simplemente hay que hacer una elección sin base, sino también que *este punto también puede estar tan cerca de la acción misma como sea posible.* ¿Por qué, en realidad buscar precisión en el segundo decimal si no estamos seguros del primero?

Esta línea de argumentación podría apoyar una interpretación decisionista de Descartes, lo que implicaría una crítica muy radical de la razón. Entonces, la única tarea de la razón sería probar el teorema de imposibilidad que acabamos de bosquejar, y a partir de tal punto, la intuición o la satisfacción estarían libradas a sus propios medios, sin ningún apoyo del razonamiento formal. Creo que esta interpretación es la menos verosímil de las tres que estoy ofreciendo sobre la filosofía moral de Descartes, porque es difícil hacerla embonar con los textos (después citados) que son el cálculo del interés propio. No obstante, parece haber una corriente en los escritos de Descartes, y de hecho en su carácter, que refleja esta preferencia por la decisión no apoyada, a la que nosotros nos adherimos rígidamen-

[51] Robbins y Stobaugh (1974), p. 130.

te, una vez tomada. Aun a falta del teorema de la imposibilidad, esta preferencia podría ser apoyada por una repugnancia aristocrática a los cálculos de cualquier índole y por una predilección no menos aristocrática por la absoluta firmeza de carácter, por muy excéntrica que sea. Es difícil leer la correspondencia de Descartes sin sorprenderse ante este aspecto de "grand seigneur" de su carácter: nunca explica, nunca ofrece disculpas.

Procedo ahora a la *tercera interpretación.* Es argüir que no sólo el *tiempo* empleado en salir de la selva aumenta si el viajero se detiene constantemente para revaluar su situación, sino que esta racionalidad instantánea en realidad hace que el *camino* mismo sea más largo de lo que habría sido siguiendo una línea recta arbitrariamente (o al menos, óptimamente) elegida. Los economistas han discutido sobre un problema muy similar en el contexto de la búsqueda del uso óptimo de la previsión en el planeamiento. Siguiendo a Milton Friedman, Clem Tisdell ha argüido, en un lenguaje muy parecido al que hemos usado aquí, que

> ... aun si el político tiene alguna capacidad (positiva) de predecir, sin embargo puede ser óptimo para él seguir una política inflexible. En realidad [este ensayo] muestra que *pasar por alto los aumentados costos de decisión* que una política "finamente sincronizada" puede implicar y los desfavorables efectos de incertidumbre que pueden generar en un grupo, las deficiencias en la capacidad de predecir pueden, aun si hay una (considerable) capacidad de predecir, hacer indeseable la política flexible o zigzagueante... surgen pautas que son útiles desde un punto de vista prescriptivo y que también explican la racionalidad de

> la conducta real que de otra manera puede parecer irracional, por ejemplo, el político que tiene una capacidad predictiva pero no se ajusta a sus predicciones a corto plazo bien puede parecer que está actuando racionalmente.[52]

En la frase que he puesto en cursivas, Tisdell distingue explícitamente su propio enfoque (correspondiente a la tercera interpretación de Descartes) del enfoque de costo de decisión (relativo a la primera interpretación). También menciona, como tercer argumento para ser inflexible, los "desfavorables efectos de incertidumbre" que después se analizan en la sección 8. Bajo ciertas condiciones (analizadas hacia el fin de esta sección), la inflexibilidad requiere un compromiso previo. Sin embargo, aquí me limitaré a repetir que el análisis de Tisdell (y tras él, el de Descartes) no implica una crítica muy radical de la razón.

Este punto se puede destacar observando uno de los principales ejemplos de Tisdell, el intercambio entre inflación y desempleo (la curva de Phillips). El problema también es analizado por William Nordhaus, en su importante obra sobre el ciclo político de los negocios. Yo diré que el enfoque de Nordhaus está más cerca de la idea de atarse a sí mismo, que por contraste no es

[52] Tisdell (1971), p. 35; las cursivas son mías. Leif Johansen (1972) ha mostrado que las cosas no son tan sencillas. La elección no se plantea entre seguir una decisión inicial como si fuera absolutamente correcta y actuar de acuerdo con sucesivas prediciones como si cada una de ellas fuese absolutamente digna de confianza; antes bien, lo óptimo es seguir un curso intermedio y tomar en cuenta hasta cierto punto las previsiones cambiantes, sin cambiar de dirección hasta el grado que lo habríamos hecho si las previsiones actuales fuesen siempre consideradas como el valor correcto.

inmediatamente pertinente en el enfoque de Descartes-Tisdell. Tisdell muestra que en ciertas condiciones, la política óptima puede consistir en mantener un nivel constante de desempleo antes que adaptarlo a las previsiones del momento. El problema de Nordhaus es completamente distinto: sus políticos no tienen que abstraerse de las previsiones actuales, sino de las consideraciones electorales. El principal punto de su análisis parece ser que votantes y planificadores siempre atribuyen cierto peso al bienestar de las generaciones futuras, mientras que los políticos, tratando de elevar al máximo el número de sus votos en la siguiente elección, sólo se preocupan por el presente. Esto tiene dos consecuencias: a largo plazo, la sociedad escogerá una política de menor desempleo y mayor inflación que la óptima; a corto plazo, cada periodo electoral empezará en austeridad y terminará en despilfarro. Entre los posibles remedios a esta tendencia en el sistema democrático, Nordhaus analiza la estrategia de Ulises:

> Una tercera posibilidad consiste en confiar la política económica a personas que no se dejarán tentar por las sirenas de la política del partido. Este procedimiento es típico de la política monetaria, que por razones históricas se encuentra centrada en los bancos centrales (como en el independiente Sistema de la Reserva Federal, en los Estados Unidos, o en el Banco de Inglaterra). Una posibilidad similar consiste en entregar la política fiscal a una Tesorería dominada por servidores civiles. Sin embargo, se podrá objetar que delegar responsabilidad a una dependencia que no es políticamente sensible a las necesidades legítimas resulta aún más peligroso que unos cuantos ciclos. Con frecuencia se plantea este peligro con respecto a los

> bancos centrales que prestan más atención a la "firmeza del dólar" o a la última chifladura monetarista que a los problemas fundamentales de la política.[53]

Volveré a tratar algunos de estos problemas en la sección 8. Aquí me limitaré a observar que la estrategia considerada corresponde a una paráfrasis de la fórmula de Pascal: "Il n'y a rien de si conforme à la politique que ce désaveu de la politique." Sacar de la esfera política la política monetaria sería, en sí, un acto político; la abdicación de los políticos, y no su destronamiento. En su obra sobre el mismo problema, Assar Lindbeck también sugiere la despolitización de algunas dependencias políticas como remedio posible.[54] Entre las otras posibles estrategias analizadas por Lindbeck, especialmente me llamó la atención su idea de que las elecciones podrían ser espaciadas al azar, de modo que los políticos no tuvieran ningún incentivo para convertir la política en un medio con fines electorales. Albert Hirschman observa que "impredecibilidad es poder";[55] podemos controlar nuestro medio, haciéndonos pasar por impredecibles. Y a la inversa, podemos controlarnos nosotros mismos haciendo que el medio sea impredecible. (*La lotería en Babilonia*, de Borges, lleva esta idea hasta su conclusión lógica, y más allá.)

Como ya lo hice notar, no creo posible decir, sobre la base de los textos, si es más verosímil la primera interpretación de Descartes o la tercera.

[53] Nordhaus (1975), p. 188.
[54] Lindbeck (1976), p. 18, n. 8.
[55] Hirschman (1977), p. 50. Un lector de la novela *The Dice Man*, de Luke Rhinehart, reconocería la fuerza —y la flaqueza— de la frase de Hirschman.

Tienen en común la crítica de la racionalidad instantánea que, en mi opinión, es el punto central que están tratando de establecer. Virtualmente no se trata de una *tentación* que haya que resistir; el obstáculo a la plena racionalidad es la ignorancia y no la flaqueza de la voluntad. La ignorancia de alguien que nunca ha aprendido geometría es distinta de la ignorancia del geómetra dormido; una vez que hemos adquirido la idea de que puede ser racional no adherirse a la racionalidad instantánea, no queda nada que nos impida seguir esa visión. Tal es, cuando menos, la conclusión preliminar que se sigue de una lectura del citado texto del *Discourse de la Méthode*, tomado en sí mismo. Citaré ahora otros pasajes que sugieren una conclusión ligeramente distinta.

En los escritos de Descartes hay una analogía interesante y (hasta donde yo sé) sumamente descuidada, entre sus ideas cosmológicas y sus opiniones sobre la sociedad. Son comunes a ambas los lemas de *atomismo* y *optimismo*; unas partículas que se desplazan erráticamente crearán por sí solas un universo perfectamente ordenado, y unos individuos motivados egoístamente actuarán de tal modo que favorezcan el bien común. Citaré dos pasajes tan similares en su redacción que es difícil no creer que tienen su fuente en alguna inspiración común:

> Pues *Dios ha establecido tan maravillosamente esas Leyes*, que aunque supusiéramos que no creó nada más que lo que ya he dicho, y aunque no pusiera en ello ningún orden ni proporción, sino que con ello compuso un Caos, el más confuso y más desordenado que los poetas puedan describir: bastan para hacer que las artes de ese Caos se desenreden unas de otras, por sí solas

y se dispongan en tan buen orden que tendrán la forma de un Mundo muy perfecto.[56]

Confieso que es muy difícil medir hasta dónde la razón nos ordena que nos interesemos por el público; pero tampoco es una cosa en la que sea necesario ser muy exacto: basta con satisfacer la conciencia y en ello se puede satisfacer mucho a la inclinación. Pues *Dios ha establecido de algún modo el orden de las cosas*, y unido a los hombres en tan estrecha sociedad, que aunque cada quien redujera todo a sí mismo y no tuviese ninguna calidad para los demás, no dejaría de emplearse ordinariamente para ellos en todo lo que estuviera en su poder, siempre que se valiera de prudencia.[57]

Las dos frases que he puesto en cursivas deben proceder del mismo molde de pensamiento. La lógica subyacente así como la elección de los términos son demasiado similares para que sea posi-

[56] Car *Dieu a si merveilleusement établi ces Lois*, qu'encore que nous supposisions qu'il ne crée rien de plus que ce que j'ai dit, et même qu'il ne mette en ceci aucun ordre ni proportion, mais qu'il en compose un Chaos, le plus confus et le plus embrouillé que les poètes puissent décrire: elles sont suffisantes pour faire que les parties de ce Chaos se émelent d'elles-mêmes, et se disposent en si bon ordre qu'elles auront la forme d'un Monde très parfait. Descartes (1897-1910), xi, p. 34.

[57] J'avoue qu'il est tres difficile de mesurer jusques où la raison nous ordonne que nous nous intéressions pour le public; mais aussi n'est ce pas une chose en quoi il soit nécessaire d'être for exact; il suffit de satisfaire a sa conscience, et on peut en cela donner beaucoup à son inclination. *Car Dieu a tellement établi l'ordre des choses*, et conjoint les hommes ensemble d'une si étroite société, qu'encore que chacun rapportât tout a soi-même, et n'eut aucune charité pour les autres, il ne laisserait pas de s'employer ordinairement pour eux en tout ce qui serait de son pouvoir, pourvu qu'il usât de prudence. *Ibid.*, IV, p. 316; las cursivas son mías.

ble otra conclusión. Veamos ahora cómo Descartes elabora la teoría sociológica implícita en el segundo de los pasajes anteriores:

> La razón que me hace creer que los que no hacen nada más que para su utilidad particular, deben, como los demás, trabajar por su prójimo y tratar de agradar a todos, mientras que esté en su poder, si quiere abusar de prudencia, y que ordinariamente se ve que quienes son considerados como oficiosos y dispuestos a causar placer reciben también gran cantidad de buenos servicios de los demás, aun de aquellos con los que no están obligados, los cuales no recibirían si les creyeran de otro humor, y que los trabajos que se toman por causar placer no son tan grandes como las comodidades que les da la amistad de quienes los conocen. Pues no se espera de nosotros más que los oficios que podemos prestar cómodamente, y no esperamos más de los otros; pero ocurre a menudo que lo que les cuesta poco nos aprovecha mucho, y aun nos puede importar en la vida. Cierto es que a veces se pierde el esfuerzo haciendo el bien, y al contrario, se gana haciendo mal; pero ello no puede cambiar la regla de prudencia, la que sólo se relaciona con las cosas que ocurren las más de las veces. Y para mí, la máxima que he observado más en toda la conducta de mi vida ha consistido en seguir solamente el gran camino, y en creer que la principal astucia es no querer emplear alguna astucia.[58]

[58] La raison qui me fait croire que ceux qui ne font rien que pour leur utilité particulière, doivent aussi bien que les autres travailler pour autrui, et tâcher de faire plaisir à un chacun, autant qu'il est en leur pouvoir, s'ils veulent user de prudence, est qu'on voit ordinairement arriver que ceux qui sont estimés officieux et prompts à faire plaisir, reçoivent aussi quantité de bons offices des autres, même de ceux qu'ils n'ont jamais obligés, lesquels ils ne recevraient pas, si on les croyait d'autre humeur, et que les peines qu'ils ont à faire plai-

Por simplicidad, podemos suponer que Descartes estaba pensando en algo como el Dilema del Preso (definido en el capítulo I, sección 4). Parece argüir entonces, que aun si es racional elegir la estrategia no cooperativa en un solo juego de este tipo, podría ser racional cooperar en una secuencia de juegos. Esta idea general puede ser más precisa de algunas maneras, algunas de las cuales se elucidan en la sección 7 del capítulo III. Aquí me limitaré a esbozar dos interpretaciones que son válidas en condiciones muy contrastantes. En primer lugar, un egoísta podría razonar de la siguiente manera: "Si exteriormente parezco una persona útil y altruista, otros acudirán a mí y se preocuparán por mi bienestar, de modo que derivarán una utilidad positiva de mi propio consumo. Esto significa que tendrán que actuar altruistamente para conmigo, y el resultado neto será que mi bienestar es mayor del que habría sido si yo hubiese actuado como un estricto egoísta; pues en general, los beneficios de un sacrificio para la persona por quien se hace excederán el costo para la persona que lo hace, de modo que dados unos sacrificios iguales en

sir, ne sont point si grandes que les commodités que leur donne l'amitié de ceux qui les connaissent. Car on n'attend de nous que les offices que nous pouvons rendre commodément, et nous n'en attendons pas davantage des autres; mais il arrive souvent que ce qui leur coûte peu, nous profite beaucoup, et même nous peut importer de la vie. Il est vrai qu'on perd quelquefois sa peine en bien faisant, et au contraire qu'on gagne a mal faire; mais cela ne peut changer la règle de prudence, laquelle ne se rapporte qu'aux choses qui arrivent le plus souvent. Et pour moi, la maxime que j'ai le plus observée en toute la conduite de ma vie, a été de suivre seulement le grand chemin, et de croire que la principale finesse est de ne vouloir point du tout user de finesse. *Ibid.*, IV, pp. 356-357.

todas partes, el beneficio de los sacrificios hechos a mi favor excederá el costo de los sacrificios que yo haga por los demás." Este argumento, que es bastante similar a las ideas recién propuestas por Gary Becker (y más estudiadas en la sección 7 del capítulo III), funciona bien si nos combinamos un solo individuo actuando al lado de otros individuos cuyas reacciones podría predecir de la manera especificada. En cambio, si se permite que los demás actores razonen del mismo modo, y sepan que todos razonan de esta manera, el argumento se desploma, porque nadie sentiría amistad hacia una persona que sólo actúa altruistamente para obtener los beneficios de la amistad.

En segundo lugar, podemos apelar al análisis de los superjuegos, propuesto por Martin Shubik y Michael Taylor.[59] Aquí, la idea básica es que si *1)* el número de juegos es infinito, o una variable aleatoria, o finito pero desconocido de los jugadores y *2)* como no se cuenta con el futuro, puede ser individualmente racional el elegir la estrategia cooperativa en el Dilema del Preso. Ambas condiciones son muy verosímiles en situaciones de la vida real. No sabemos *cuándo* moriremos (de modo que se satisface la primera condición), pero sí sabemos *que* moriremos (y por tanto, se satisface la segunda). Sin embargo, el truco está en que la conducta cooperativa no es una estrategia dominante en el superjuego, aunque sea parte de la solución. Esto significa que una persona no escogería esa estrategia a menos que estuviese muy segura de que los otros harían lo mismo. En otras palabras, el segundo enfoque a la "racionalidad del altruismo" tropieza con el problema opuesto al primero: el segundo

[59] *Véase*, en particular, M. Taylor (1976a).

enfoque sólo triunfa si es público y notorio de todos, el primero sólo si es oculto e insospechado. Esto también implica que en un mundo con incertidumbre y desconfianza —el mundo real—, *ninguno de los dos* enfoques triunfará.

Sin embargo, apartémonos de este problema y supongamos que Descartes tiene razón al argüir que a largo plazo es racional ayudar a otros aun si es irracional a corto plazo. Volviendo ahora a la comparación entre Descartes y Pascal, podemos preguntar cómo proceder si hemos decidido que *"le grand chemin"* es el curso óptimo. ¿Bastarán los cálculos abstractos antes sugeridos para hacer que se nos preste ayuda, aun cuando sea a expensas del ayudador, o antes bien deberá tratar éste de establecer un carácter del cual acuda ayuda "sans violence, sans art, sans argument"? ¿Debe la razón estar presente en cada elección, o sólo de una vez por todas en la elección que fija el *hexis* a partir del cual se seguirán las elecciones posteriores "automáticamente"? Pascal dice que este último procedimiento es el mejor, y yo creo que tiene razón. Los argumentos antes esbozados no tienen una naturaleza tan intuitiva y obligatoria que sea posible mantenerlos ante mí cada vez que me vea yo tentado a mi conveniencia a corto plazo. El mismo argumento sigue siendo válido para el viajero que trata de orientarse para salir de la selva. Aun si ha decidido no cambiar de dirección *"pour de faibles raisons"*, estas razones pueden parecer más apremiantes cuando la sed y la fatiga abrumen. El actor metarracional, que puede predecir esta posibilidad, se pondría entonces unas "anteojeras" de modo que le fuese físicamente imposible recabar la información que preste cierta fuerza a estas razones; una analogía visual de la estrategia

de taparse las orejas con cera, elegida por Ulises para sus hombres. Una analogía a la estrategia elegida por Ulises para sí mismo sería prefijarse un curso y luego, de alguna manera, hacerse incapaz de cambiarlo. De hecho, ninguna de estas estrategias corresponde al análisis cartesiano de *Les passions de l'âme*, que, en cambio, recomendaría un cambio planeado de carácter, de modo que aun si alguien pudiese recabar la información y actuar en consecuencia, no querría hacerlo. (Para un análisis de estas tres técnicas de atarse a sí mismo, *véase* la sección 9.)

5. Preferencias incongruentes de tiempo

Este problema fue planteado por primera vez por R. H. Strotz hace unos 25 años, en un artículo justamente célebre. Está relacionado de modo directo con el problema del cambio endógeno de preferencias: el tema de la sección 6. Para ver la importancia de estos problemas y la relación existente entre ellos, empecemos por considerar la habitual noción de racionalidad. De acuerdo con la definición común, un actor racional sencillamente es uno que tiene preferencias congruentes y completas *desde cualquier punto en el tiempo*. Creo yo que la noción de hombre racional debe extenderse para incluir consideraciones temporales. Para ser precisos, algunos requerimientos de conciencia deben imponerse, a la vez, a la *elección de sucesiones* del actor y a su *sucesion de elecciones*. La no satisfacción del primer requerimiento queda ejemplificada por preferencias incongruentes de tiempo, y del segundo, por un cambio endógeno de preferencias.

Como ya se ha dicho, ambos problemas están

íntimamente relacionados, y es posible tratarlos dentro de un marco unificado, como en un reciente artículo de Peter Hammond.[60] Sin embargo, yo creo que hay buenas razones para seguir la tradición, y tratarlos por separado. Tanto las preferencias incongruentes de tiempo como las preferencias endógenamente cambiantes indican que no se siguen las decisiones pasadas, y ambos mecanismos presuponen que esto puede ocurrir aun en ausencia de influencias externas como persuasión, propaganda, discusión o experiencia. En este sentido, ambos mecanismos son no-exógenos. Sin embargo, unas preferencias incongruentes de tiempo no implican ninguna modificación de carácter del individuo, como sí lo hacen los cambios endógenos de preferencia. No cabe duda de que estos cambios, siendo endógenos, deben explicarse por medio de algunas leyes fijas de transformación que caractericen al individuo y que pueden servir como rasgos constantes de carácter de orden superior. El cambio de carácter es simplemente el desenvolvimiento de un potencial. La elección hecha de acuerdo con preferencias y constantes de tiempo, por otra parte, es el desenvolvimiento progresivo de una actitud *real* e inmutable hacia el tiempo.

En general es aceptado (pero muy lejos de ser-

[60] Hammond (1976). Arguye en la p. 163 que el compromiso previo "en realidad no es un modo de resolver la incongruencia", y que "no significa más que determinar lo que realmente es factible en cualquier etapa". Luego se retracta un tanto, lo bastante para nuestros propósitos, añadiendo que "semejante ejercicio a menudo está lejos de ser trivial; en realidad, con frecuencia es como una especie de progreso técnico". Yo no tengo nada contra esta caracterización: la planeación congruente es sagacidad dentro de los límites de lo factible, el compromiso previo es sagacidad equivalente a una modificación de los límites.

lo universalmente) que para un individuo, el simple hecho de tener preferencias de tiempo, por encima de lo que queda justificado por el hecho de que somos mortales, es irracional y tal vez, asimismo, inmoral.[61] (Para las *sociedades* que planearan por un futuro infinito podría ser irracional *no* tener preferencias de tiempo, porque de otra manera, el consumo se aplazaría indefinidamente, en bien de mayor inversión.)[62] Una opinión común al respecto es, quizá, la siguiente: como seres racionales y dejando a un lado el problema de que moriremos pero no sabemos

[61] Esto es enérgicamente afirmado por John Rawls (1971, pp. 284 *ss.)* para el caso intergeneracional (pero *véase* n. 62, *infra)* y por Thomas Nagel (1970) para el caso intrageneracional. Argumentos ingeniosos, en pro de la opinión contraria, han sido propuestos por Derek Parfit (1973) y por B. A. O. Williams (1976). Por el argumento, adoptaré aquí la opinión de Rawls-Nagel, de modo que las preferencias de tiempo siempre constituyen un *problema*. Sin embargo, debo añadir que hay mucho por decir en favor de la idea de Parfit-Williams, pero puede suponerse que la mayor parte de mis argumentos serían aceptados aun por ellos, para los casos (y sin duda hay tales casos) en que las preferencias de tiempo corresponden a flaqueza de voluntad y no (como en los casos analizados por Williams) a los rasgos de carácter que dan sustancia a mi vida. También debe añadirse que los argumentos contra el altruismo indiscriminado (tratar a cualquier desconocido tan altruistamente como a mi mujer) son más fuertes que los argumentos en contra del tratamiento indiscriminado de todos los momentos temporales; el primero parece menos compatible con la sustancia de carácter que el segundo.

[62] Si una sociedad tratara de elegir la tasa de ahorro que maximiza el total de consumo sin descuento sobre un tiempo infinito, descubriría que no hay tal tasa de ahorro: por cualquier tasa inferior a 100 % hay una tasa superior que da un consumo total mayor, pero la propia tasa de 100 % da consumo cero. (Esta afirmación, desde luego, depende de cierto número de suposiciones, explicadas, por ejemplo, en Heal [1973].)

cuándo, deseamos distribuir nuestro bienestar parejamente a lo largo del tiempo. Un año es un año en un año; no hay motivos racionales para preferir el presente al futuro simplemente porque es presente. (Supongo aquí que las preferencias de tiempo implican siempre una preferencia por el presente ante el futuro, lo que pasa por alto el interesantísimo caso del avaro y los importantes "casos mixtos" donde el hombre es un avaro para volverse después un despilfarrador.) Por otra parte, está el hecho de la falta de voluntad. Sencillamente, no siempre podemos seguir nuestras inclinaciones racionales, y las preferencias de tiempo pueden considerarse como la expresión formal de la incapacidad. Alguien que prefiera el presente sobre el futuro *deberá* asignar mayores cantidades de consumo al presente que al futuro, sin hacer caso a los argumentos racionales o complicados (*véase infra*) en favor de lo contrario. Esta interpretación de las preferencias del tiempo como flaqueza de voluntades es decisiva para mi argumento, que no se aplica a los casos (si existen) en que las preferencias se deben a un juicio considerado de que ciertas partes de la vida son, inherentemente, más valiosas que otras.[63]

El tener una preferencia por el presente me conducirá, característicamente, a *lamentar* mis decisiones pasadas cuando surge un nuevo presente —el pasado futuro— y se revela candidato

[63] Peter Hammond ha sugerido (en conversación) que podríamos decidir gastar más ahora y menos en la vejez, porque en el periodo posterior el consumo inferido quedaría compensado por los felices recuerdos del periodo anterior. Muchas objeciones podrían plantearse contra esta idea; aquí, sólo mencionaré dos. Primera, ¿no podría ser el consumo menor compensado *ahora* por gratas previsiones de una riqueza ulterior? Segunda, las exter-

igualmente digno para mi atención. Ahora bien, si alguien lamenta una decisión pasada, lo racional será *reconsiderarla.* Si en el pasado, yo decidí hacer una asignación de bienes a (lo que entonces era) el futuro, y si llego a lamentar la decisión ¿no sería lo racional reconsiderar la decisión para lo que me queda del futuro? El problema es que mi único desacuerdo con mi decisión pasada puede ser con la asignación de bienes entre el pasado y el presente, mientras que mis egos anterior y actual pueden convenir en lo que hoy es el futuro. Y, en realidad, los economistas no han argüido que reconsiderar una decisión pasada sea un modo de hacerla más racional. Se ha dicho, antes bien, que tal evaluación representa un elemento adicional de irracionalidad, al menos en las condiciones que ahora exploraremos para precisarlas más plenamente.

Suponemos que un individuo tiene una unidad de un bien de consumo que será distribuida aparentemente sobre cierto número de años. En los ejemplos numéricos *infra,* sólo consideraremos los casos de tres años y de cuatro años. Los "años" serán interpretados como unidades de tiempo "naturalmente voluminosas", que reflejan una división no arbitraria del continuo temporal. Por razones que analizo más adelante, considero importante ver el presente como algo más que un punto de distancia cero a partir de ahora. Es, fenomenológicamente, el periodo que incluye el presente matemático y que está separado de

nalidades temporales del consumo podrían ser negativas y no positivas, de modo que en mi vejez yo podría recordar con amargura mis anteriores goces. Al optimismo de Tennyson, "mejor haber amado y perdido / Que no haber amado nunca", podemos oponer el pesimismo de Donne, "Mejor ser feo que haber sido hermoso". El problema exige ya un tratamiento matemático.

alguna manera "natural" de otros periodos. Tales divisiones naturales incluyen los días, que están separados entre sí por periodos de sueño "cuando no existo". El tiempo que va de un día de pago a otro día de pago ("empezaré a ahorrar el próximo día de pago"); un año calendárico ("Dejaré de beber el 1 de enero del año próximo"), etcétera. En cualquier año dado y las preferencias de tiempo hasta el año pasado (n) del periodo sobre el cual fueron asignados los fines pueden resumirse en una función de utilidad $U = u_i\ (C_1,\ C_2,\ \ldots\ C_n)$. Nos limitamos a funciones de utilidad de la forma general

$$u_1 = a\ln C_1 + b\ln C_2 + c\ln C_3 + d\ln C_4$$
$$u_2 = a\ln C_2 + b\ln C_3 + c\ln C_4$$
$$u_3 = a\ln C_3 + b\ln C_4$$

con la modificación obvia para el caso de tres periodos. Estas funciones logarítmicas de utilidad son matemáticamente convenientes, y tienen sentido económico por el rasgo de reducir la utilidad marginal. Desde luego, hay una infinidad de otras funciones con la misma característica, pero dado que la lógica del argumento propuesto abajo es existencial y no universal, basada en contraejemplos no en teoremas generales, esta restricción realmente no importa.

Deben notarse otras dos características de estas funciones de utilidad. En primer lugar, suponemos que son constantes, en el sentido de que en el año i la importancia relativa del consumo en el año i comparada con el consumo en el año $i+1$ es la misma que la importancia relativa en el año j del consumo en el año j comparada con el consumo en el año $j+1$, para todo i y j hasta $n-1$, inclusive. (Esta es una afirmación precisa de lo que se dijo antes, que analiza-

remos el problema de las preferencias de tiempo independientemente del problema de las preferencias cambiantes.) En segundo lugar, las funciones logarítmicas de utilidad permiten una evaluación sencilla de la racionalidad de una asignación dada. Para una asignación dada (C_1, C_2, C_3, C_4), en que $\Sigma_i C_i = 1$, definimos el *valor racional* de la asignación como $a\ln C_1 + a\ln C_2 + a\ln C_3 + a\ln C_4$, lo cual significa que todos los años reciben igual peso. El candidato obvio para un índice de racionalidad sería entonces la proporción de este valor racional al valor racional de la asignación racional (1/4, 1/4, 1/4, 1/4). Sin embargo, como sólo estamos interesados en comparaciones ordinales, podemos simplificar más las cosas y definir el índice de racionalidad como

$$r(C_1, C_2, C_3) = 27 \cdot C_1 C_2 C_3$$
$$r(C_1, C_2, C_3, C_4) = 256 \cdot C_1 C_2 C_3 C_4$$

para los casos de tres periodos y de cuatro periodos, respectivamente.

Dados estos preliminares, nos encontramos en posición de definir y de analizar la noción de preferencias *congruentes* de tiempo. Empezaremos con algunos ejemplos de tres periodos. Al comienzo del primer año el individuo escoge una asignación que maximiza $u_1(C_1, C_2, C_3)$, sujeto a la limitación presupuestaria $\Sigma_i C_i = 1$. Supongamos que esto produce los valores C'_1, C'_2, C'_3. Al comienzo del segundo año, el individuo planea una asignación que maximiza $u_2(C_2, C_3)$, sujeta a $C_2 + C_3 = 1 - C'_1$. Supongamos que esto produce los valores C''_2 y C''_3. La congruencia de las preferencias de tiempo pueden quedar entonces definidas como el requerimiento de que $C'_2 = C''_2$; que la cantidad planteada en el primer año

para el segundo año es la misma que la cantidad elegida en el segundo año, cuando llega. La definición puede extenderse fácilmente al caso del periodo $-n$. Con funciones de utilidad de la forma dada, es fácil mostrar que las preferencias de tiempo son congruentes si y sólo si $a/b = b/c$ en el caso de los tres periodos, y $a/b = b/c = c/d$ en el caso de cuatro periodos. En otras palabras, la constancia y la congruencia de las preferencias del tiempo implican una pérdida exponencial del futuro. (Este es un resultado general, que no depende de la forma específica de la función de utilidad.) Para un ejemplo numérico tomemos

$$u_1 = 4 \ln C_1 + 2 \ln C_2 + \ln C_3$$
$$u_2 = 4 \ln C_2 + 2 \ln C_3$$

lo que nos da la asignación (4/7, 2/7, 1/7). Esta asignación es *irracional,* en el sentido de que muestra una preferencia con el presente, pero, no obstante, es *congruente,* ya que no hay necesidad de revisiones de planes anteriores. Por tanto lo llamaré la *asignación congruentemente irracional.*

Como nuestro primer ejemplo de preferencias inconsistentes de tiempo, tomemos:

$$u_1 = 3 \ln C_1 + 2 \ln C_2 + \ln C_3$$
$$u_2 = 3 \ln C_2 + 2 \ln C_3$$

Una persona que mostrara estas preferencias planearía en el año uno, la asignación (1/2, 1/3, 1/6), pero cuando llegara en realidad el segundo año, se reconsiderará la asignación para los dos últimos años, convirtiéndose en (3/10, 2/10). En total, esto significa que la *asignación incongruentemente irracional,* será (5/10, 3/10, 2/10). La

persona incongruentemente irracional no sólo es miope, sino miope en forma inherentemente contradictoria, que nunca se atreve a apegarse a sus decisiones pasadas. Strotz preguntó si tal persona —a pesar de su incongruencia y su irracionalidad— no podría volverse lo bastante racional para comprender su situación y hacerle frente de manera racional. Esto no es tan extraño como puede parecer. Consideremos a una persona que ha heredado una fortuna y decidido gastarla de la manera siguiente: en el primer año, gastará la mitad de su fortuna en una enorme parranda, y luego dividirá la otra mitad parejamente sobre el resto de su vida. Sin embargo, al llegar el segundo año, prefiere —fiel a sus preferencias (incongruentes, pero constantes)— gastar la mitad de la mitad restante en una parranda un tanto más pequeña en el segundo año, luego dividir lo que le queda sobre el resto de su vida, y así sucesivamente. (Obsérvese que el punto no es que se ha vuelto aficionado a la *dolce vita*; analizamos esta posibilidad en la sección 6.) No resulta psicológicamente inverosímil, que esa persona, doblemente irracional —que tiene preferencias de tiempo, pero incongruentes— pueda conocer o llegar a conocer su propio carácter y tomar precauciones contra sus egos anteriores, para que no lo traicionen. En este caso, la solución obvia es comprarse un seguro anual, que no puede reconvertir en efectivo. Es fácil ver que un seguro anual no sirve contra la irracionalidad *congruente*. O bien la persona compra un seguro que disminuye exponencialmente con el tiempo, pero entonces, en todo caso, hace lo que él de todos modos habría hecho por sí mismo (suponiendo congruencia). O bien compra un seguro que le deja la misma cantidad cada año:

pero entonces, es racional. O bien se compra un seguro anual que le deja una cantidad mayor en el primer año y cantidades menores (iguales o exponencialmente más reducidas) en cada año ulterior... pero este es el caso de la incongruencia irracional.

La irracionalidad incongruente puede ser un fenómeno más frecuente e importante que la irracionalidad congruente. E. S. Phelps y R. A. Pollack han propuesto un mecanismo general para explicar cómo las preferencias de tiempo se vuelven no exponenciales y por ello incongruentes.[64] (Su obra trata de la planificación intergeneracional, pero llega a cubrir el caso general; en realidad, el enfoque parece más apropiado para este último.) Arguyen que las preferencias de tiempo pueden descomponerse en dos funciones distintas de descuento: una que da una prioridad absoluta al presente sobre todos los tiempos posteriores, que reciben pesos menores e iguales; y una que se reduce en forma exponencial. Yo considero esto psicológicamente sano, porque en la idea de preferencia de tiempo unimos dos problemas distintos que pueden y deben ser separados: la prioridad absoluta del presente, y la gradual disminución del futuro. La prioridad absoluta del presente es algo como mi prioridad absoluta sobre otras personas: yo soy yo, mientras que ellos, todos están "ahí". La disminución es un fenómeno de perspectiva que admite grados de "ahí fuera": el futuro remoto es como un pariente lejano, mientras que el futuro cercano se parece a un pariente directo. (En este punto se requiere la noción de un presente "denso", pues en el tiempo continuo sería difícil dar sentido a la prioridad absoluta del presente (ma-

[64] Phelps y Pollack (1968).

temático).) De estas dos funciones de descuento la primera, por sí misma o en conjunto con la segunda, destruye la congruencia de las preferencias. Esto ofrece un incentivo al individuo que se ata a sí mismo; mientras se siente incapaz de contener sus deseos actuales, sí se siente capaz de establecer alguna maquinaria que le impedirá sucumbir a los deseos futuros que no hayan aún alcanzado este *status* absoluto. La lógica de la situación está resumida en la plegaria de San Agustín: "Dame castidad y continencia, pero todavía no." [65]

En el ejemplo numérico de preferencias incongruentes, antes definido, podemos definir la *asignación de Ulises* como (1/2, 1/3, 1/6). Es decir, la persona es capaz de atarse a sí misma, a la asignación que inicialmente prefiere, para lograr congruencia. Ahora bien, la tesis general de este capítulo es que tales estrategias de compromiso previo son las "segundas mejores" soluciones que forman el núcleo de la teoría de la racionalidad imperfecta, pero este caso parece ofrecernos un contraejemplo. De hecho, observamos que el índice de racionalidad de la asignación de Ulises es 3/4, mientras que el de la asignación incongruentemente irracional es 81/100. La persona ingenua y miopemente incongruente logra una asignación más uniforme (es decir, más racional) sobre el tiempo, que la persona sagaz, capaz de atarse a sí misma. Si la persona sagaz llegara a comprender esto, ¿no sería la suprema sagacidad renunciar a la sagacidad ("la principale finesse est de ne vouloir point du tout user de finesse"), y limitarse a seguir su propensión natu-

[65] *Confesiones*, VIII, vii. Davidson (1969) llamó mi atención a este pasaje.

ral, aun si esto implicara una reconsideración de las decisiones pasadas?

Antes de responder a esta pregunta, será útil mirar otro enfoque refinado: la *estrategia de la planificación congruente*, como fue analizada por Strotz y más elaborada por R. A. Pollak.[66] Argüiré que este enfoque es de dudosa validez, pese a lo muy ingenioso. Supongamos ahora que la persona sagaz, incapaz de atarse o renuente a hacerlo, razona de esta manera: "Si durante el primer año escojo cierta cantidad C_1 (no importa cómo haga esta elección), entonces tendré $1 - C_1$ disponible para consumo en años posteriores. A falta de compromiso previo y dado mi carácter tal y como lo conozco, asignaré esta cantidad a los dos últimos años, de acuerdo con u_2. Esto significa que asignaré $(a/(a+b))\,(1-C_1)$ al segundo año y $(b/(a+b))\,(1-C_1)$ al tercero. Esto otra vez implica que sólo podrán realizarse secuencias de consumo de la forma general

$$\left(C_1,\ \frac{a}{a+b}(1-C_1),\ \frac{b}{a+b}(1-C_1)\right).$$

Sabiendo esto, elegiré la mejor entre estas secuencias, es decir, elegiré C_1 para que maximice

$$\left(a \ln C_1 + b \ln\left(\frac{a}{a+b}(1-C_1)\right) + c \ln\left(\frac{b}{a+b}(1-C_1)\right)\right)$$

[66] Pollak (1968). Su análisis emplea el tiempo continuo, y mi argumento es justamente una elaboración de un caso de tiempo discreto de tres periodos, ofrecido en una nota en la p. 203 de su escrito.

Para el caso antes analizado ($a = 3$, $b = 2$, $c = 1$), la consiguiente, *asignación de planificación congruente* es (18/53, 21/53, 14/53), con un índice de racionalidad de cerca de 0.96.

Este es un resultado sorprendente. La asignación de planificación congruente resulta muy cercana a la asignación de planificación racional, incluyendo un poco más en el primer año, muchísimo más en el segundo y considerablemente menos en el tercero. Sin embargo, nótese que aun el tercer año resulta mejor en la estrategia de planificación congruente que en la asignación incongruentemente irracional y en la asignación de Ulises. Hay una paradoja en esta racionalidad de planificación congruente, pues si alguien pudiera (es decir, tuviese la fuerza de voluntad) de seguir esta estrategia, entonces sin duda también podría seguir la estrategia racional. Si la solución es factible, entonces el problema no existe. A esta objeción podría responderse con el siguiente contraargumento. La estrategia de la planificación congruente es una respuesta al problema de las preferencias *incongruentemente* irracionales, no al problema de las preferencias incongruentemente *irracionales*. La cuestión es cómo enfrentar a mi predecible falta de capacidad de actuar según mis intenciones actuales, y no enfrentarme al componente irracional de estas mismas intenciones. Esta línea de argumento podría utilizarse, asimismo, para responder al acertijo antes planteado con respecto a la estrategia de Ulises, sobre si la ingenuidad podría ser más sagaz que la sagacidad. La estrategia de Ulises es una precaución contra la incongruencia, no contra la irracionalidad; de hecho, logra la congruencia al costo de apartarse más aún de la irracionalidad, al menos en el ejemplo dado.

Antes de proponer una respuesta al contraargumento, será útil tener una mayor variedad de estructuras a la mano. En el cuadro II.1 de la página 129 enumero algunas funciones de utilidad y las asignaciones (con los índices de racionalidad) que se siguen para cada una de ellas, incluyendo, para facilitar la comparación, los casos antes analizados.

Antes de comentar el cuadro, los casos E y F requieren unas cuantas palabras como explicación.

Cuando pasamos del caso de tres periodos al de cuatro, tenemos una elección entre dos interpretaciones de la estrategia de la planificación concongruente. ¿Debemos interpretar esta estrategia como una sagaz decisión, en el año uno, tratando de obtener el máximo provecho de las ingenuas decisiones que predicen para los años dos, tres o cuatro, o debemos imputar al tomador de decisiones del año uno la suposición de que la decisión del año dos también será una estrategia (de bajo nivel) de planificación congruente? Pollack opta por la última necesidad, y (como lo observa en otro contexto)[67] en realidad es razonable suponer; *1)* que la sagacidad, una vez adquirida, nunca se perderá, y *2)* que un sagaz tomador de decisiones tiene conciencia de *1)*

El caso G es el ejemplo de las preferencias congruentes antes empleado. El caso A es el ejemplo numérico de las preferencias incongruentes empleadas en el análisis preliminar. Los casos B, D y F corresponden todos ellos, al caso de la gran parranda durante el primer año, y luego la división pareja, que antes analizamos informalmente. Los casos C y F se añaden para tener alguna variedad adicional. Observamos que en to-

[67] Phelps y Pollack (1968).

CUADRO II.1.

		Asignación racional	*Asignación de Ulises*	*Asignación de planificación congruente*	*Asignación de planificación incongruente*
A	$u_1 = 3 \ln C_1 + 2 \ln C_2 + \ln C_3$ $u_2 = 3 \ln C_2 + 2 \ln C_3$	1/3, 1/3, 1/3; $r = 1$	1/2, 1/3, 1/6; $r = 0.75$	18/53, 21/53, 14/53; $r = 0.96$	1/2, 3/10, 1/5; $r = 0.81$
B	$u_1 = 3 \ln C_1 + \ln C_2 + \ln C_3$ $u_2 = 3 \ln C_2 + \ln C_3$	1/3, 1/3, 1/3; $r = 1$	3/5, 1/5, 1/5; $r = 0.65$	9/25, 12/25, 4/25; $r = 0.75$	3/5, 6/25, 4/25; $r = 0.62$
C	$u_1 = 4 \ln C_1 + 3 \ln C_2 + 2 \ln C_3$ $u_2 = 4 \ln C_2 + 3 \ln C_3$	1/3, 1/3, 1/3; $r = 1$	4/9, 1/3, 2/9; $r = 0.89$	48/167, 68/167, 51/167; $r = 0.97$	4/9, 20/63, 5/21; $r = 0.91$
D	$u_1 = 4 \ln C_1 + 2 \ln C_2 + 2 \ln C_3$ $u_2 = 4 \ln C_2 + 2 \ln C_3$	1/3, 1/3, 1/3; $r = 1$	1/2, 1/4, 1/4; $r = 0.84$	4/13, 6/13, 3/13; $r = 0.88$	1/2, 1/3, 1/6; $r = 0.75$
E	$u_1 = 3 \ln C_1 + 2 \ln C_2 + \ln C_3$ $u_2 = 3 \ln C_2 + 2 \ln C_3 + \ln C_4$ $u_3 = 3 \ln C_3 + 2 \ln C_4$	1/4, 1/4, 1/4, 1/4; $r = 1$	1/2, 1/3, 1/6, 0; $r = 0$	96/361, 90/361, 105/361, 70/361; $r = 0.96$	1/2, 1/4, 3/20, 1/10; $r = 0.48$
F	$u_1 = 2 \ln C_1 + \ln C_2 + \ln C_3 + \ln C_4$ $u_2 = 2 \ln C_2 + \ln C_3 + \ln C_4$ $u_3 = 2 \ln C_3 + \ln C_4$	1/4, 1/4, 1/4, 1/4; $r = 1$	2/5, 1/5, 1/5, 1/5; $r = 0.819$	8/47, 12/47, 18/47, 9/47; $r = 0.816$	2/5, 3/10, 1/5, 1/10; $r = 0.61$
G	$u_1 = 4 \ln C_1 + 2 \ln C_2 + \ln C_3$ $u_2 = 4 \ln C_2 + 2 \ln C_3$	1/3, 1/3, 1/3; $r = 1$	4/7, 2/7, 1/7; $r = 0.63$	4/7, 2/7, 1/7; $r = 0.63$	4/7, 2/7, 1/7; $r = 0.63$

dos los casos, salvo en el caso F, la estrategia de planificación congruente es la más racional. La asignación de Ulises es la más racional en el caso F. La asignación de Ulises es a veces menos racional que la asignación incongruentemente irracional, y a veces es más racional. Cuando es la más racional, la congruencia y la racionalidad van de la mano; de otra forma, son opuestas entre sí. Esto al menos, se sostiene si excluimos la estrategia de la planificación congruente, como creo que debemos hacerlo. Procedo ahora a ofrecer mi argumento contra la utilidad y hasta contra el sentido de la estrategia de la planificación congruente.

El argumento más poderoso nos lo ofrece un examen de los casos C y F. Nos enfrentamos aquí a un tomador de decisiones con preferencias claramente marcadas por el presente, lo que en nuestra interpretación de las preferencias del tiempo significa, sencillamente, que no tiene la suficiente fuerza de voluntad para alcanzar la racionalidad. Y sin embargo, se nos pide creer que semejante persona podría adoptar la estrategia de la planificación congruente, lo que en estos casos implica *menos* consumo en el primer año —el año de la decisión— que en todos los años siguientes. Es como si el incongruente despilfarrador pudiese resolver su caso de incongruencia convirtiéndose en un avaro; como si las preferencias incongruentes para el presente sobre el futuro pudiesen volverse congruentes, viéndolas en preferencia por el futuro sobre el presente. Creo yo que el punto es éste: una persona carente de fuerza de voluntad así como de congruencia puede decidir una estrategia de planificación congruente para minimizar el daño que sufrirá por obra de sus futuros egos, pero su falta de fuerza de voluntad le impedirá llevar adelante tal decisión. (*Cf.* el análisis

de la norma iv en la sección 2.) La estrategia de la planificación congruente podría ser una solución al problema de la incongruencia sólo si no fuera más racional (es decir, si no requiriera más fuerza de voluntad) que la asignación incongruentemente irracional, pero éste no es el caso en ninguno de los ejemplos del cuadro.

Yo puedo imaginar una respuesta a esta respuesta al contraargumento, a saber, que la asignación del año 1 que está fuera del alcance de mi fuerza de voluntad cuando se le describe como parte de un plan racional puede parecer más atractiva cuando se le *redescribe* como parte de una solución de "paquetes" que hasta cierto punto toma en cuenta a cuáles preferencias u_i. Esta sugerencia tendría un parecido de familia, con la idea de Ainslie de apuestas privadas indirectas, que también gira sobre una redefinición o una redisposición de mi espacio interno. Es decir, la sugerencia sería que para algunas personas puede ser imposible actuar racionalmente si el motivo para la conducta racional es actuar racionalmente, mientras que pueden alcanzar la racionalidad si fue obtenida como subproducto de acciones emprendidas para lograr la congruencia misma. Esta idea parece extraña, pero no, tal vez, completamente inverosímil. Como ya se observó en la sección 2 hay casos en que simultáneamente deseamos alcanzar x y sin embargo no deseamos emprender ninguna acción *para* alcanzar x.

Creo haber mostrado que, con la posible excepción analizada en el párrafo anterior, la estrategia de la planificación congruente no es una opción factible ante la incongruencia y la irracionalidad. En algunos casos, la asignación de Ulises puede lograr a la vez congruencia y la mejor aproximación a la racionalidad, pero en otros casos logra

la congruencia a expensas de la racionalidad. La elección entre la asignación de Ulises y la asignación incongruentemente irracional depende, entonces, de cuál deseo es más fuerte: el deseo de congruencia (el deseo de imponer mi decisión a mis egos posteriores) y el deseo de racionalidad (mi preocupación por el bienestar de mis egos futuros). Creo yo que en tales casos de conflicto moral, la razón debe prevalecer, por lo que la hipersagaz decisión de renunciar a la sagacidad sería la correcta. Atarse a sí mismo favorece la congruencia, pero el precio pagado puede ser excesivo. Situaciones como A, C y E en el cuadro son auténticas excepciones a la tesis de que la mejor manera de alcanzar la racionalidad imperfecta es por medio del compromiso previo; desde luego, a menos que estemos dispuestos a investigar las preferencias mismas. Ahora, estudiaremos este caso.

6. El cambio endógeno de preferencias

Una teoría general de la acción humana —explorada más en el capítulo III— puede esbozarse de la manera siguiente. Para explicar por qué una persona en una situación dada se comporta de un modo y no de otro, podemos ver su acción como resultado de dos sucesivos procesos de filtración. El primero tiene el efecto de limitar el conjunto de acciones posibles en abstracto al conjunto *factible,* es decir, el conjunto de acciones que, simultáneamente satisfacen cierto número de limitaciones físicas, técnicas, económicas y político-legales. El segundo tiene el efecto de seleccionar un miembro del conjunto factible como la acción que se emprenderá. A partir de esto vemos inmediatamente que el acto de atarse a sí mismo

puede ir dirigido hacia un cambio del conjunto factible o hacia un cambio del mecanismo que elige la alternativa factible que va a realizarse. En la sección 5 supusimos que la estrategia de Ulises implicaba un cambio inducido en el conjunto factible; aquí exploraremos la segunda posibilidad. Supongamos que en cualquier momento dado de tiempo, la selección de un miembro del conjunto factible se efectúa por medio de una elección racional (para otros mecanismos del segundo filtro *véase* la sección 6). Esto implica que el compromiso previo debe ocurrir actuando sobre las preferencias. Esto puede tomar la forma de intentar deliberadamente producir un cambio en la estructura de preferencia, o resistir deliberadamente a un cambio que de otra manera habría sido causado por mecanismos endógenos. Nos interesará principalmente la última posibilidad.

Este problema está virtualmente inexplorado en la bibliografía. Hay un cuerpo bibliográfico pequeño pero que está creciendo rápidamente, sobre la idea de las preferencias endógenamente cambiantes, pero muy rara vez se supone que un agente puede adoptar una actitud activa o estratégica hacia sus propias preferencias.[68] Antes bien, se le

[68] Los análisis más recientes y profundos se encuentran en el simposio sobre hábitos formados, *Journal of Economy Theory* 13 (1976). Las primeras contribuciones incluyen Haavelmo (1944, pp. 17 *ss.)*, Georgescu-Roegen (1950), Gorman (1967), Peston (1967), Pollack (1970), Von Weiszäcker (1971), Cyert y De Groot (1975). Una nota escéptica suena en Stigler y Becker (1977). El enfoque estratégico es utilizado por Shefrin y Thaler (1977) y por March (1977). Los primeros vuelven a plantear el problema de Strotz de tal manera que se vuelve el cambio planeado de preferencias, pero sus postulados psicológicos subyacentes, con una distinción entre "planificadores" y "actores" parecen bastante endebles. El enfoque elegido por

ve como el pasivo vehículo de preferencias de acuerdo con cierta lógica que él mismo no comprende. Algunos autores mencionan, de paso, que la persona puede tomar sus precauciones contra tal cambio si llega a comprender el mecanismo por el cual se produce. La resistencia a las drogas que causan hábito es el ejemplo habitual. Sin embargo, las implicaciones de este enfoque rara vez son elaboradas, y el mayor interés se fundamenta en los aspectos analíticos y manipulativos: ¿cómo cambian (endógenamente) las preferencias, y cómo puede esto ser explotado por otras personas? Analizaré estos aspectos y luego los relacionaré con nuestro tema principal.

Empezaré con un análisis del seminal escrito de C. C. von Weiszäcker sobre los cambios endógenos de gustos.[69] Por simplicidad, presupone *1)* que las preferencias actuales sólo dependen del consumo en el periodo inmediatamente anterior, y *2)* que el ingreso del consumidor será asignado sólo a dos bienes. Obras recientes han mostrado [70] que la segunda suposición es fundamental para los resultados obtenidos por Von Weiszäcker y que no generalizan hasta el caso del artículo-*n*. Aquí no haré caso a este problema, ya que creo que hay cuestiones de importancia filosófica que tienen que analizarse hasta en el caso especial. La primera suposición es un poco sorprendente pues

March está muy cercano por el adoptado aquí, haciendo hincapié en las preferencias incongruentes y endógenamente cambiantes.

[69] Von Weiszäcker (1971). "La endogenación de variables extraeconómicas" podría emplearse como caracterización general de la labor más interesante que está efectuándose en la economía contemporánea. Von Weiszäcker es un maestro en este arte; *véase* también Von Weiszäcker (1973).

[70] Pollack (1976).

excluye la noción de formación acumulativa del carácter, que muchos escritores tenían en mente cuando analizaron los cambios endógenos de preferencia.[71] Sin embargo, es probable que la introducción de toda la historia del consumo como variable complicara el análisis hasta hacerlo virtualmente insostenible, y que en la actual etapa de investigación, la estrategia elegida por Von Weiszäcker sea muy sabia.

Los economistas, cuando estudian la elección de los consumidores, utilizan más o menos intercambiablemente tres nociones distintas: preferencias, funciones de utilidad y funciones de demanda. Dadas varias condiciones, que supondremos satisfechas, cualesquiera dos de estos conceptos pueden deducirse o construirse a partir del tercero. Supongamos que el consumidor, en el tiempo t, tiene una estructura de preferencia P_t, representada por una función de utilidad u_t, de la cual podemos derivar una función de demanda f_t la cual correlaciona con cada conjunto de precios e ingresos cierto vector de artículos (de dos componentes) escogido por el consumidor: $q_t = f_t(p, m)$. Suponemos entonces que la demanda actual es una función de la demanda y el consumo en el periodo inmediatamente anterior: $f_t = F(f_t - 1, q_t - 1)$, donde F es alguna funcional que representa los rasgos de carácter constante subyacentes en las preferencias cambiantes y que gobiernan su cambio. Más específicamente, suponemos que F es tal que existe una g tal que $q_t = g(p, m, q_t - 1)$. Esta última forma es la única utilizada por Von Weiszäcker. Aplicaciones sucesivas de la función g generan una sucesión g_i de grupos de artícu-

[71] *Véase* especialmente Georgescu-Roegen (1950). La excesiva generalidad de este escrito puede explicar su esterilidad en las investigaciones posteriores.

los, que en ciertas condiciones (cuya interpretación económica no es clara) convergen en un punto $q = h(p, m)$ que es independiente del conjunto inicial. La relación h puede ser interpretada formalmente como una función de demanda, en cuanto relaciona cada conjunto de precio-ingreso con cierto grupo de artículos. A esta función de demanda también corresponde (dadas ciertas suposiciones discutidas) una función de la utilidad U, "como si", en el sentido de que podemos predecir el grupo $q = h(p, m)$ suponiendo que el consumidor está maximizando U, dadas p y m. Por último, podemos construir la "estructura de preferencia a largo plazo" P con la propiedad xPy si y sólo si $U(x) > U(y)$.

Retornemos a las preferencias instantáneas P_t. En realidad no se necesita una expresión deíctica por tiempo, porque la única diferencia pertinente entre P_{t_1} y P_{t_2} es el consumo en los periodos inmediatamente anteriores $t_1 - 1$ y $t_2 - 1$. Si el consumo en estos periodos fuese el mismo, entonces también serían las mismas las preferencias en los tres periodos. Por consiguiente, podemos hacer expresiones deícticas más útilmente de preferencias por consumo en el periodo anterior, escribiendo $xP(y)z$ pues, "x es preferida a z, dado que el consumo en el periodo inmediatamente anterior fue y". Ahora Von Weiszäcker probó el siguiente teorema, válido en condiciones que verbalmente pueden expresarse como inercia o conservadurismo del consumidor:

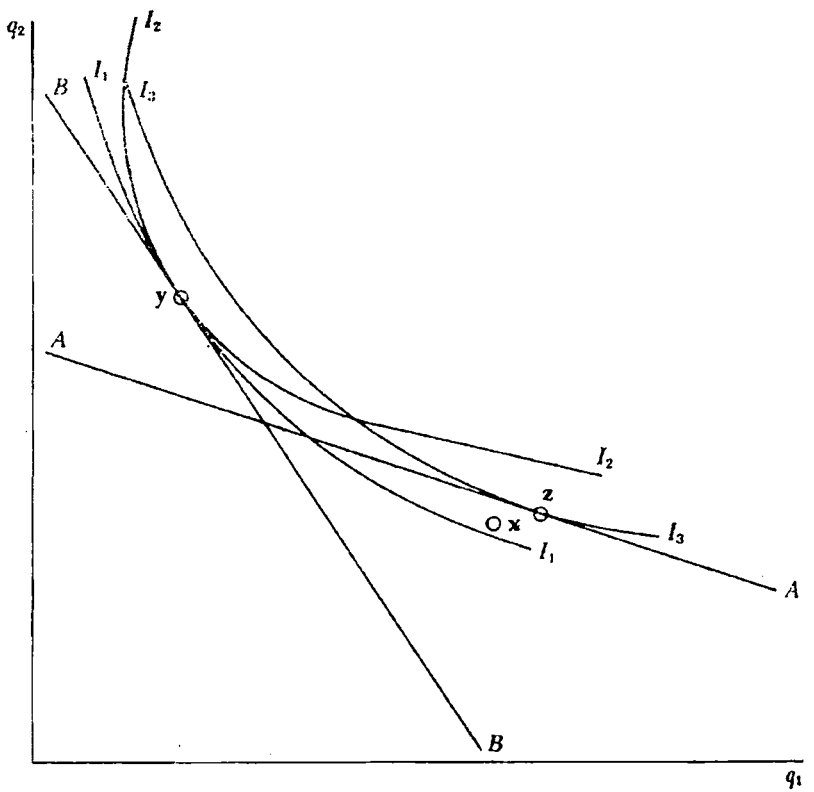
> Dados dos vectores de artículos x y y, entonces xPy si y sólo si existe una sucesión de vectores $r_1 \ldots r_n$ tal que $r_1P(y)y$, $r_2P(r_1)r_1$, $r_3P(r_2)r_2 \ldots r_n P(r_{n-1})r_{n-1}xP(r_n)r_n$.

El teorema dice, en otras palabras, que x es mejor que y de acuerdo con mis preferencias a largo plazo si y sólo si existe un camino de y a x, tal que cada paso del camino sea una mejora de acuerdo con las preferencias instantáneas, al comienzo del paso. *Si* las preferencias a largo plazo son interpretadas como mis preferencias "reales" o mis preferencias "racionales", entonces el teorema implica que aun si hay un conflicto entre mis preferencias a corto plazo y mis preferencias a largo plazo, en el sentido de que $yP(y)x$, este conflicto podrá siempre resolverse mediante el uso de estrategias indirectas. Es como una fantasía infantil hecha realidad: un paseo por un paisaje en que desde cualquier punto siempre se puede subir a cualquier punto más alto sin tener que ir cuesta arriba, porque el panorama contornea el cambio, como función de nuestro camino.

Von Weiszäcker también ofrece un ejemplo ingenioso (y discutible) de este enfoque en acción. En la gráfica II.1 las cantidades de bienes q_1 y q_2 son medidas a lo largo de los ejes. La línea AA representa la línea del presupuesto de un granjero. La línea BB representa la línea del presupuesto de un obrero de la industria. Inicialmente, la persona en cuestión es un granjero que está disfrutando del conjunto de consumo y. I_1I_1 es la curva de indiferencia a largo plazo que pasa por y, en tanto que I_2I_2 es la curva de indiferencia a corto plazo correspondiente a las preferencias en y. El punto z es el punto de equilibrio a largo plazo si el granjero se muda a la ciudad. El punto x representa un conjunto de consumo que tiene las propiedades siguientes:

(1) zPx
(2) xPy
(3) $yP(y)x$

Si el granjero consume x mientras vive en la ciudad, quedará un excedente que el fisco le gravará. Ahora bien, la vida en la ciudad (tanto en la versión z como *a fortiori* en la versión x) no parece muy atractiva desde el punto de vista del granjero, aun cuando (en ambas versiones) sea superior de acuerdo con sus preferencias a



GRÁFICA II.1.

largo plazo. Ahora, el teorema nos dice que existe una secuencia de puntos r_i que puede llevarlo, sin dolor, de y a x. (La razón de que consideremos la transición a x y no a z quedará en claro dentro de un momento.) Sin embargo, el truco está en que, al trazar las curvas, la parte inicial de la sucesión debe encontrarse por encima de ambas líneas del presupuesto, porque todas las r_i deben estar por encima de I_1I_1. Esta última afirmación se justifica como sigue. Por definición, cada r_1 satisface la condición del teorema de que exista una sucesión de mejoras a corto plazo, de x a r_i (a saber, r_i, $r_2 \ldots r_i - 1$), y de acuerdo con el teorema esto implica que r_iPy para todas r_i, lo que una vez más significa que todas las r_i deben encontrarse en curvas de indiferencia más altas que I_1I_1. Von Weizsäcker propone la siguiente solución a este problema. El gobierno podría subsidiar los puntos intermedios r_i que se encuentran por encima de AA. Cuando llegue una r_i por encima de AA, el gobierno transferirá al granjero a la ciudad, le dará un ingreso que le permita comprar x y conservar la diferencia entre z y x para sí, con objeto de recuperar el dinero gastado en subsidios. El granjero se encuentra mejor (por causa de xPy) y el gobierno tiene una ganancia porque los subsidios sólo se dan por periodo limitado, mientras que el ingreso por impuestos puede perpetuarse indefinidamente, por medio de un gravamen a los descendientes del granjero.[72] Así, todo es para bien en este mundo con optimalidad de Pareto.

¿De veras? Estamos analizando el caso en que una agencia externa podría *manipular* al indivi-

[72] En realidad, Von Weizsäcker (1971) sólo muestra que esto es verdad si la tasa social de descuento es lo bastante baja.

duo con gustos endógenamente cambiantes, colocándolos en una sucesión de situaciones que le llevarán con seguridad y sin dolor a cierto estado que no habría podido alcanzar por sí solo. Si se cree que esto sea éticamente aceptable, deberemos suponer que el individuo —de ser racional— habría hecho lo mismo por sí solo, dado el mismo conocimiento del proceso causal subyacente en el cambio de preferencia. Es decir, deberemos suponer que habría solicitado un préstamo para colocar los puntos intermedios dentro del conjunto alcanzable, planeando pagar el préstamo cuando llegara al estado final. No niego que esto puede ocurrir en algunos casos. En particular, creo que la elección de una educación universitaria debe verse en esta perspectiva. Por otra parte, soy incapaz de ver que fuera irracional y un individuo rechazara la oferta de un subsidio, o se negara a solicitar un préstamo. Y si pudiera ser racional rechazar tales ofertas, entonces supongo que es antiético manipular al individuo por medio de tales ofertas.

Este punto merece entrar en ciertos detalles. Distingamos entre coerción, seducción, persuasión y elección voluntaria. Las diferencias entre estos conceptos surgen si consideramos la relación entre las preferencias antes del hecho y las preferencias después del hecho. La coerción y el compromiso voluntario son los dos extremos de la gama. La *coerción* ocurre cuando una persona prefiere x sobre y, y continúa haciéndolo aun cuando alguien lo fuerce (físicamente) a hacer y. *Elección voluntaria* significa que la persona inicialmente prefiere y sobre x, y hace y por esa razón. (Tal vez llevara a preferir x y y después del hecho, pero esto no importa en un análisis de las razones para hacer y.) La *seduc-*

ción ocurre cuando una persona inicialmente prefiere x sobre y, pero llega a preferir y sobre x después de ser obligado a hacer y. La seducción de Cecile de Volanges por Valmont es la clásica excepción literaria de este proceso. Amartya Sen, al analizar ese problema, planteado por Von Weiszäcker, parece creer que la seducción es éticamente justificable:

> El labrador en cuestión podía preferir, en realidad, estar en un sector y no en otro, es decir, trabajando en agricultura campesina y no en empleo asalariado en la ciudad. Absurdo sería atribuir todas estas preferencias a "irracionalidad" de cualquier índole; puede haber algo que no sea ni remotamente irracional en preferir ser su propio amo en su propia granja, aun con menores ingresos, en lugar de desempeñarse como "esclavo del salario" en una fábrica, o en realidad, en alguna granja. Sin embargo, más fácil de defender es el tratar de evitar la suposición de gustos estacionarios. Las preferencias por nuestro modo de vida y nuestra ubicación son, típicamente, el resultado de nuestras experiencias pasadas, y una renuencia inicial a cambiar no implica un disgusto perpetuo. Esta distinción tiene cierto efecto sobre los aspectos de beneficencia de la política de empleos, ya que la importancia que uno desea atribuir a la brecha entre salarios como reflejo de la preferencia del trabajador tenderá a depender del grado en que se espera que los gustos aumenten como consecuencia del propio movimiento.[73]

El mismo argumento, transferido a *Les liaisons dangéreuses*, probablemente no encontraría muchos partidarios. Ahora bien, Von Weiszäcker está estableciendo un argumento más poderoso para justificar el traslado de personas del cam-

[73] Sen (1975), pp. 53-54.

po a la ciudad, dependiendo de la persuasión más que de la seducción. Así pues, *persuasión*, significa que un individuo es llevado por una sucesión de mejoras a corto plazo, a preferir *y* sobre *x*, aun si inicialmente prefería *x* sobre *y*. Mi afirmación es que la persuasión es más similar a la seducción que a la elección voluntaria. Esto, sin embargo, se sostiene cuando la persuasión no va acompañada por una declaración de intenciones, que informe al individuo de que está a punto de ser manipulado. Nunca podrá justificarse el explotar mecanismos intrapsíquicos que son desconocidos de la persona. Esto, desde luego, no implica que la persona deba ser protegida de las influencias del medio que tiendan a cambiar sus preferencias; el aprendizaje y la experiencia, "el camino anguloso" [the angled road],[74] son esenciales para el desarrollo de la persona; en cambio, sí implica que planear la experiencia de otra persona va contra la ética; también implica que debemos rechazar la afirmación de Robert Nozick de que las seducciones nunca son coercitivas;[75] lo son si se hacen con el propósito de conducir a la persona a un lugar al que acaso no quisiera ir. No hay diferencia esencial entre coerción y seducción, ni entre seducción y esta forma de persuasión.

[74] Experience is the Angled Road
Preferred against the Mind
By — Paradox — the Mind itself —
Presuming it to lead.

Quite Opposite —How Complicate
The Discipline of Man—
Compelling Him to Choose Himself
His Preappointed Pain.

(Dickinson, 1970, núm. 910).

[75] Nozick (1969a).

Aristóteles tiene una respuesta a este argumento. "Sólo una persona totalmente insensata puede dejar de saber que nuestros caracteres son el resultado de nuestra conducta; pero si un hombre actúa a sabiendas en una forma que dé por resultado el volverse injusto, debe decirse que es voluntariamente injusto." [76] En otras palabras, la incapacidad del individuo para predecir la repercusión de su conducta actual sobre preferencias futuras es señal de su irracionalidad y, por implicación, no debe formar la base para culpar al inductor. Esta observación parece correcta en el caso común en que el carácter (*hexis*) formado por una serie de acciones de cierta índole es precisamente la disposición a efectuar ese tipo de acciones. Hoy, al menos, ciertamente debemos saber que el fumar producirá la adicción a fumar. Sin embargo, el funcionamiento de la mente también está sujeto a leyes más complejas. La repetición de cierta acción puede producir una disposición a *no* efectuar ese tipo de acción, como cuando una persona se satura de caramelos para provocar una aversión a los dulces; y puede ser difícil saber de antemano, qué actos de repetición forman hábito y cuáles curan del hábito. El mecanismo puede ser aún más complejo, como cuando una serie de acciones *A* interactúan con una disposición existente *B* para producir una tendencia a efectuar las acciones *C*. (¿Conduce el uso de la mariguana al de la heroína?) Resulta irrazonable exigir que todos tengan un conocimiento cabal de estos mecanismos, e igualmente irrazonable pasar por alto el hecho de que muchos grupos tienen un incentivo para adquirir información acerca de ellos y explotarlos. En otras palabras, *nadie* puede adqui-

[76] *Etica Nicomaquea 1114.*

rir el conocimiento acerca de *todas* estas tentaciones que un *grupo* especializado podría adquirir acerca de *una* de ellas. La orden "Conócete a ti mismo" en realidad sería lógicamente incongruente, porque los procesos causales que implica son, a su vez, rasgos de carácter (de orden superior) que pueden ser modificados por la conducta, y así sucesivamente.[77]

Al analizar la persuasión, debemos distinguir entre dos cuestiones. Una de ellas es si es moralmente obligatorio que el persuasor informe al objeto de la persuasión acerca de sus intenciones; esta pregunta ha sido contestada en sentido positivo. Una segunda es si puede ser racional que el individuo rechace la oferta. Si se responde a esta pregunta en sentido negativo, la primera pierde gran parte de su interés. Sin embargo, yo creo que hay que responder en el sentido positivo. Como lo ha observado Robert Pollack,[78] el programa por el cual un "no fumador miope es llevado a convertirse en un gran fumador... es enteramente congruente con las suposiciones de Von Weiszäcker", y sin embargo es difícil concluir que el individuo se encuentra mejor en el estado final que en el estado inicial. El paso que va de preferencias "a largo plazo" o "estables" a preferencias "reales" o "racionales" no fue justificado por Von Weiszäcker. Nada es tan estable y tan definitivo como la muerte.

El análisis del Estado paternalista puede verse a la luz de estas ideas. En la medida en que la prohibición a anunciar cigarrillos se debe a los propios consumidores, reales o potenciales, que desean protegerse contra las Sirenas de la publicidad, no podemos hablar de paternalismo.

[77] *Cf.* también *infra*, cap. II, sec. 9.
[78] Pollack (1976).

Lo mismo puede decirse de la propuesta a introducir un día a la semana, libre de televisión, para estimular las actividades políticas o culturales; o el empleo obligatorio de los cinturones de seguridad. Huelga decir, creo yo, que las personas son libres de atarse a sí mismas (por medio de leyes), a atarse a sí mismas (con cinturones de seguridad) o a proteger sus valores "más profundos" contra los más impulsivos. Los problemas arduos surgen cuando estas restricciones son impuestas a las personas contra sus preferencias *ex ante facto* (tal vez *ex post facto*). Me parece bastante claro que el gobierno no tiene el derecho de proteger al público contra la televisión un día de cada siete si el público no desea ser protegido. Es mucho menos obvio que el gobierno no deba tener el derecho de hacer obligatorio el empleo de cinturones de seguridad, pues aquí se está pidiendo al Estado pagar la cuenta si ocurren accidentes.

La pregunta, en pocas palabras, es si el Estado de bienestar implica o justifica el Estado paternalista. Algunos dicen sí, y su aversión al paternalismo les hace rechazar el Estado de bienestar. Otros también dicen sí, y aceptan el paternalismo porque aceptan el Estado de beneficencia. Personalmente, yo tiendo a decir no, y a argüir por un Estado de bienestar sin paternalismo. Si la gente no desea atarse a sí misma, no se le debe obligar a elegir entre estar atada y no recibir ayuda si se perjudica como resultado de no estar atada. Participación espontánea en actividades riesgosas o, aun de mayor importancia, la negativa espontánea a participar en actividades de educación física (¿qué decir acerca del ejercicio físico obligatorio durante la hora de la comida, para ahorrar dinero de hospitales al Esta-

do?), es algo que hay que respetar. (Aquí, desde luego, es decisivo que si hay alguna correlación entre la medida en que las personas se dedican a actividades riesgosas y su contribución a los fondos de beneficencia social: de otra manera, su negativa a atarse a sí misma no sería una expresión de espontaneidad, sino un intento de aprovechar lo ganado por otros.) A la postre, el problema se establece entre una vida larga y saludable y una vida rica de acuerdo con las preferencias de cada quien. Si un grupo de individuos bien informados decide que prefiere tener menos recursos para hacer lo que les gusta y más recursos para hacer lo que les han impuesto, ¿qué autoridad podría anular su decisión?

7. El compromiso previo en la conducta animal

Entre las muchas sorprendentes observaciones que nos ofrece Ainslie se encuentra la prueba de que los animales son capaces de emplear estrategias de compromiso previo. Dos de estos ejemplos son los siguientes:

> Es bien sabido que las palomas picotean una tecla que haya sido asociada con alimento, aun cuando no picotear produzca una recompensa mucho mayor. Ainslie descubrió que las palomas picotearían regularmente una tecla roja para tener acceso a alimento, durante dos segundos, mientras que no picotear les produciría acceso a alimento, durante cuatro segundos, empezando tres segundos después. Si la tecla se encendía con luz verde, doce segundos antes de que debiera encenderse de rojo, algunos de los sujetos acudían a picotearla, las más de las veces, cuando su único

efecto era impedir que la clave se encendiera de rojo más avanzada la prueba. Al parecer, los sujetos aprendieron a picotear la tecla cuando estaba en verde sólo si esto anulaba la eficacia temporal de la clave roja.[79]

La tendencia de las palomas hambrientas a evitar activamente ciertas oportunidades de obtener alimento parece incongruente con el concepto ortodoxo de recompensa. Si el número fijo de picoteo a una tecla requerida para una sola recompensa de alimento es alto pero no tan alto que el picoteo cese (entre 50 y 2000), las palomas picotearán una segunda tecla cuyo único efecto es hacer que la tecla original sea inaccesible durante un periodo ("tiempo fuera")... los sujetos que buscaban "tiempo fuera" de la oportunidad de obtener alimento, fueron los que, como puede suponerse, se enfrentaron a elecciones seguidas entre descansar y trabajar por una pobre recompensa. En semejante situación se habría esperado que su elección vacilara entre las alternativas. Si suponemos que una conducta ambivalente es menos grata que la de aferrarse a cualquier alternativa, los sujetos pueden buscar cualquier recurso que ate su futura relación a uno u otro, o al menos que reduzcan la frecuencia del cambio.[80]

El primer caso es un ejemplo directo de flaqueza de voluntad, superada mediante el método del compromiso previo. El segundo caso también tiene claras analogías en el comportamiento humano; podemos recordar, por ejemplo, al cartesiano viajero en el bosque. Estos descubrimientos, al parecer contradicen una implicación importante de esta obra, a saber, que los hombres no son ángeles (es decir, plenamente racionales) ni animales (es decir, en esencia miopes); son criaturas

[79] Ainslie (1975), pp. 472-473.
[80] *Ibid.*, p. 476.

imperfectamente racionales, capaces de enfrentarse estratégicamente a su propia miopía. ¿Implican los resultados de que informa Ainslie que los hombres deben ser rebajados al nivel de animales, o los animales elevados a la categoría de hombres? Para presentar mi argumento en favor de una respuesta negativa a esta pregunta, estableceré ciertas distinciones desarrolladas en el capítulo I, sección 3 y más elaboradas en el capítulo III, sección 7. En el capítulo I, sección 3 se dijo *1)* que la capacidad de maximización global es rasgo exclusivo y distintivo del hombre, y *2)* que los ejemplos de maximización global (es decir, el empleo de estrategias indirectas, como una intersección en el camino, en la conducta animal) descubiertos en el reino animal pueden evolucionar sólo por accidente. El hombre tiene una capacidad *general* de maximización global, que puede mostrar en situaciones cualitativamente nuevas y sin precedente, mientras que los escasos ejemplos de tal conducta en animales suelen ser específicos de cada situación.

En este capítulo, defiendo una forma más compleja de maximización global, que incluye el uso de estrategias indirectas (es decir, compromiso previo) con el propósito de permitir unas estrategias indirectas (es decir, inversión). La capacidad de una conducta imperfectamente racional en este sentido también es general; la estrategia de atarse a sí mismo se aplica a una gran variedad de situaciones y puede explotar una gama sumamente vasta de mecanismos. El hecho de que pueda acondicionarse a las palomas a atarse a sí mismas en situaciones que son a la vez muy artificiales y muy específicas no debe interpretarse como prueba de un empleo espontáneo del compromiso previo en una gama muy

vasta de medios naturales. Y aun si pudiésemos demostrar que los animales emplean la conducta del compromiso previo en situaciones de la vida real, la explicación de esto habría que buscarla en mecanismos evolutivos, y cómo preadaptación, preliotropia o deriva genética. El atractivo de una capacidad generalizada para bastarse a sí mismo debe reservarse al hombre. En particular, es engañoso el empleo del término "buscá" por Ainslie en la última frase del segundo párrafo antes citado.

Repitiendo algunas de las ideas desarrolladas en el capítulo I, existe en la conducta humana una *presunción de racionalidad* tal que la maximización global o las estrategias del compromiso previo —la mejor, o la "segunda mejor" racionalidad— son formas naturales de conducta que no requieren ninguna explicación. Antes bien, procede una explicación cuando se observan desviaciones de la racionalidad perfecta e imperfecta. En el reino animal, el mecanismo general de la selección natural crea una *presunción de miopía*, y cualquier caso dado de sacrificio a corto plazo o conducta de compromiso previo precisaría una explicación separada. En la medida en que puede mostrarse que los animales se comportan espontáneamente en una de estas maneras orientadas hacia el futuro, deberemos atribuirles experiencias mentales en el mismo sentido en que los seres humanos tienen experiencias mentales. La capacidad generalizada para una maximización global (no estereotipada) o una conducta de compromiso previo presupone un espacio interior en que están presentados los posibles estados. Para que las consecuencias futuras establezcan una diferencia en la elección presente, de alguna manera deben estar presentes en la con-

ciencia *y estar presentes como no realizadas y simplemente posibles;* supongo que esto es testimonio de una vida mental en un sentido muy fuerte. Tales experiencias mentales pueden encontrarse en los vertebrados superiores, pero las palomas de Ainslie carecen del elemento de espontaneidad requerido.

8. Abdicación del poder

La situación de Ulises también interviene en toda discusión sobre política. Aquí, trataré de dos casos que parecen de importancia especial: el problema de la democracia y el problema del Estado capitalista.

Una democracia directa —ya sea en el sentido de que todos los ciudadanos votan sobre todas las cuestiones en lugar de elegir representantes, en el sentido de que se puede suspender a los representantes en cualquier momento— tenderá a una política zigzagueante y a una constante reva luación de los planes pasados; será incontinente, vacilante e ineficaz. Hegel sólo es uno de los muchos filósofos de la política que han argüido que esta libertad total se convierte a una total ilibertad, tanto en el sentido conceptual de que la libertad ilimitada *es* esclavitud como en el sentido causal de que la anarquía de la libertad total tiende a allanar el terreno a un dictador. Indicaré brevemente cómo las democracias clásicas y modernas han desarrollado diferentes maneras de enfrentarse a este problema.

En la Atenas clásica, todas las decisiones importantes se tomaban en la asamblea de ciudadanos, que se reunían al menos 40 veces por año. La asamblea podía reconsiderar en cualquier

momento sus decisiones anteriores, lo que habría hecho de ella una fácil presa de demagogos o de otros manipuladores si no se hubiesen establecido ciertas salvaguardias institucionales. M. I. Finley ha llamado la atención hacia dos de estas instituciones. La primera es el ostracismo, que en la práctica era el derecho de expulsar a los demagogos conocidos. En contraste con Ulises, la asamblea no siempre podía impedirse actuar siguiendo el canto de las Sirenas (pero *véase* la segunda institución, *infra);* sin embargo, sí podía evitar ese canto, expulsando a los cantores. La segunda institución, era llamada *grafe paranomon,* "por la que un hombre puede ser acusado y juzgado por hacer una 'propuesta ilegal en la Asamblea'... *aun si tal propuesta hubiese sido aprobada por la Asamblea*".[81] Finley observa que esta disposición difiere de las correspondientes salvaguardias de las democracias modernas: "Allí estaba la lógica en el *grafe paranomon,* en la noción de que este procedimiento, el demos estaba echando una segunda ojeada a la propuesta, y no que una rama del gobierno, el poder judicial, le estaba revisando la acción de otra rama, la legislatura." [82] También sugiere un contraste entre el enfoque griego a la estabilización de la política y los científicos políticos que han argüido que cierto grado de apatía es requisito necesario para la viabilidad de la democracia.[83] Yo afirmaré ahora que estas dos observaciones, aunque parcialmente correctas, son engañosas en aspectos importantes.

En primer lugar, Finley parece haber subestimado la sutileza del *grafe paranomon* (como lo

[81] Finley (1973), p. 26.
[82] *Ibid.*, p. 80.
[83] *Ibid.*, p. 67.

describió él mismo). Según su análisis, la institución podía compararse a las leyes, recién aprobadas en algunos países, que permiten a los consumidores anular contratos de pagos en abonos, durante cierto periodo. (Strotz arguye que comprar en abonos es un recurso para el compromiso previo,[84] pero como en el caso maniaco-depresivo de la sección 2 del presente capítulo, puede ser razonable comprometernos previamente contra los compromisos previos.) Sin embargo, yo creo que una comparación mejor sería si el vendedor no sólo corriera el riesgo de que anularan su contrato, sino que en realidad se expusiera una multa, en caso de tal anulación. Todo el que desee explotar la impulsividad de la gente sabrá entonces que corre el riesgo de ser castigado, por un impulso posterior, o cuando el impulso queda ante la razón, y esto debió de actuar como disuasivo contra tal explotación. El *grafe paranomon* era un recurso estabilizador que no sólo permitía a la razón recuperar el predominio, sino también para reducir la probabilidad de que empezara por perder ante la pasión.

En segundo lugar, las instituciones griegas y el fenómeno de apatía difícilmente pueden aparecer como mecanismos estabilizadores. Creo yo que es absurdo sugerir que una persona optaría deliberadamente por la apatía para protegerse contra su tendencia a la excesiva intromisión. La apatía puede tener o no tener este efecto; puede tener —como se explicó en el capítulo I, sección 5 — o mucho más probablemente puede no tener esta *función;* pero nunca pudo haber sido organizada con esta *intención.* Es imposible comprender una analogía colectiva con la autoinducida pérdida pascaliana de la razón. Por otra

[84] Strotz (1955-1956), p. 178.

parte, las instituciones del ostracismo y de *grafe paranomon* bien pudieron ser establecidas para mantener la democracia dentro de los límites requeridos por la eficiencia y la estabilidad. Y de acuerdo con la norma *i)* de la sección 2 del presente capítulo, esto se requiere para las instituciones que caen bajo el rubro de compromiso previo. El argumento mismo de que tienen este mismo efecto estabilizador o función (sugestión mucho más oscura) no bastaría para esta caracterización.

En las democracias modernas, cierto número de instituciones puede ser interpretado como recurso para el compromiso previo. Ya observé en la sección 4 del presente capítulo que el banco central puede considerarse como un depósito de razón contra las pretensiones a corto plazo de la pasión, argumento que será descrito con cierto detalle por Francis Sejersted.[85] Para la política liberalista clásica, la tasa de cambio y el nivel de precios eran parámetros fundamentales que en ninguna circunstancia debían transformarse en variables del control político. El mercantilismo (y el actual neomercantilismo) han subrayado la necesidad de unas adaptaciones actuales al medio cambiante: ¿se habría mandado atar Ulises al mástil si sabía que las aguas poco profundas, en torno de la isla de las sirenas, eran demasiado difíciles para que las dominara cualquiera, salvo él? (El lector recordará que este problema fue planteado en la cita de William Nordhaus, en la sección 4 del presente capítulo.) Otras instituciones a las que les ha asignado una autonomía similar y por razones similares incluyen los ministerios del exterior de muchos paí-

[85] Sejersted (1973); también Elster (1975b).

ses y el modelo de transmisiones de la BBC, en contraposición con el modelo de ORTF. Para estas instituciones es posible identificar, con varios grados de precisión, el acto de abdicación por el cual los políticos han decidido que ciertos valores son demasiado importantes o ciertas armas son demasiado peligrosas para quedar sometidas al actual control de los políticos. La supresión de la política monetaria, la política exterior o las transmisiones desde la esfera política es, en sí misma, un acto político.

En otro nivel, el sistema de elección periódica puede interpretarse en la misma perspectiva, especialmente cuando el gobierno no tiene la capacidad de disolver el parlamento y ordenar nuevas elecciones. (La desconfianza de los plebiscitos es una expresión de la misma actitud.) Según esta interpretación, las elecciones periódicas son el método que tienen los electores, de atarse y de protegerse contra su propia impulsividad. Sin embargo, ya observamos que para los *políticos*, el sistema de elecciones periódicas constituye una tentación permanente de unir todas las medidas impopulares al comienzo del periodo electoral y las populares hacia el fin, con consecuencias subóptimas. También hemos tomado la sugerencia de Lindbeck de que los políticos en turno pueden atarse ellos mismos para evitar esta tentación, mediante unas elecciones espaciadas al azar. Esto aportaría una nueva respuesta al problema de quién debe guardar a los guardianes, y en realidad las elecciones espaciadas al azar podrían ser la defensa de los electores para, a la vez, atarse ellos mismos y a los políticos. Este recurso al azar, atractivo en algunos aspectos, sería demasiado costoso en otros para que merezca ser de consideración. (Objeciones similares se apli-

can a los métodos probabilistas de elección colectiva.)[86]

En la sección 4 del presente capítulo hemos analizado dos argumentos para utilizar reglas, en lugar de la discreción al formular la economía política: el argumento del costo de la información y lo que podríamos llamar el argumento del costo de adaptación (ofrecido por Friedman y Tisdell). Aquí, consideraremos dos argumentos adicionales: uno de ellos relacionado con la suboptimalidad creada por consideraciones estratégicas, y otro relacionado con los efectos desfavorables de la incertidumbre en materia de incentivos. El argumento de la suboptimalidad ha sido recién discutido por Finn Kydland y Edward Prescott,[87] en un marco bastante similar al que estamos empleando aquí. Su principal idea es que la teoría del control óptimo, que en cualquier momento del tiempo selecciona la mejor decisión y dado que las decisiones serán similarmente seleccionadas en el futuro, no es el arma más apropiada para la planificación económica:

> Las razones de este resultado no intuitivo son las siguientes: la teoría del control óptimo es un recurso apropiado de planificación para situaciones en que los resultados actuales y el movimiento del sistema de Estado sólo dependen de decisiones de política actual y pasada, y del Estado actual. Pero no es probable que este sea el caso

[86] La elección social por un mecanismo aleatorizante, por ejemplo, dejar la probabilidad de que alguna opción sea escogida igual a la proporción de votantes que la tienen como primera elección, tiene la doble ventaja de garantizar los derechos de la minoría y de ser "a prueba de estrategias". Por otra parte, el método podría ser explicado para que la elección social fuese demasiado fácilmente reversible.

[87] Kydland y Prescott (1977).

> para sistemas económicos dinámicos. Las decisiones actuales de agentes económicos dependen en parte de sus expectativas de futuras acciones políticas. Sólo si estas expectativas fuesen invariantes con respecto a la política futura seleccionada sería apropiada la teoría del control óptimo.[88] ... Así, parece probable que la práctica actual de seleccionar la política que es mejor, dada la situación actual, convergerá en la política congruente pero *subóptimas*... La razón de que tales políticas sean subóptimas no se debe a miopía. Se toma en consideración el efecto de esta decisión sobre todo el futuro. Antes bien, la suboptimalidad surge porque no hay mecanismo que induzca a los políticos *futuros* a tomar en consideración el efecto de su política, por vía del mecanismo de expectación, sobre las decisiones *actuales* de los agentes.[89]

El problema se relaciona con dificultades tratadas en otra parte por Phelps y Pollack [90] y por Kelvin Lancaster,[91] pero Kydland y Prescott dan una nueva dimensión al problema, introduciendo la distinción entre reglas y decisiones discrecionales. Afirman que en muchos casos, la suboptimalidad generada por las actuales decisiones discrecionales puede ser eliminada por el uso de reglas fijas que impiden consideraciones estratégicas de la índole analizada. El siguiente es un ejemplo interesante, relacionado con el análisis del capitalismo como juego diferencial entre trabajadores y capitalistas, hecho por Lancaster:

> Un grupo mayoritario, digamos los trabajadores que controlan la política racionalmente, podrían

[88] *Ibid.*, p. 474.
[89] *Ibid.*, p. 481.
[90] Phelps y Pollack (1968).
[91] Lancaster (1973).

> decidir tener una constitución que limite su poder, es decir, de expropiar la riqueza de la clase capitalista. Los que tienen las menores tasas de descuento ahorrarán más si saben que su riqueza no será expropiada en el futuro, aumentando así el producto marginal y por tanto los salarios y reduciendo el precio del capital.[92]

El argumento puede leerse en un sentido débil y en otro fuerte. En un sentido débil, sólo dice que *dado* el hecho de que las decisiones de invertir están en manos de la clase capitalista, simplemente es sensato de parte de la clase obrera garantizar a los propietarios del capital una recompensa sobre sus inversiones. En un sentido fuerte (implicado por la referencia a la diferencia de tasas de descuento entre obreros y capitalistas) dice que va en interés de la clase obrera dejar las decisiones de inversión a los propietarios de capital, porque de otro modo, los trabajadores preferirían el consumo hoy al consumo después e invertirían demasiado poco. Este último argumento ya es antiguo, lo que a menudo se ha empleado como justificación del capitalismo y de las

[92] Kydland y Prescott (1977), p. 486. Es bastante sorprendente que los autores, en este respecto (o en realidad, en su artículo) no se refieran a la obra de Lancaster (1973). Si la frase "expropiar la riqueza de la clase capitalista" es interpretada en el sentido de una expropiación de las *ganancias* (y no de los medios de producción y del poder sobre las decisiones de invertir), entonces su problema es idéntico al planteado por Lancaster, y el análisis de éste puede suponerse que debió de tener cierta pertinencia para ellos. En particular debiéramos preguntar si la simetría del juego diferencial entre obreros y capitalistas no debe llevarnos a esperar que la clase capitalista también considere útil atarse a sí misma; podríamos esperar que el efecto de este doble compromiso previo pudiese alcanzar el óptimo alcanzable por medio del actual control de las variables.

desigualdades de ingreso. El primero es más original, por la introducción del elemento estratégico y la metaestrategia (compromiso previo) propuesta para superar la suboptimalidad estratégica.

Volviéndonos ahora al problema de incentivo creado por las incertidumbres, están cercanamente relacionados los argumentos en pro del "gobierno de la ley" planteados por Max Weber y muchos otros que han subrayado la importancia de un marco institucional y estable para una economía creciente. El problema a menudo se plantea como la necesidad de poder *predecir* las decisiones de la burocracia, pero el argumento de Kydland-Prescott muestra que esto es una simplificación. La capacidad de predecir es condición necesaria pero no suficiente para una optimalidad dinámica. Sin embargo, a falta de esta condición, pueden surgir consecuencias funestas. Las incertidumbres acerca de las decisiones futuras pueden deberse a dos fuentes distintas: la ley puede ser *ambigua* (como en el antiguo derecho chino)[93] o constantemente (e impredeciblemente) *cambiante* (como en el neomercantilismo contemporáneo). Muchos autores, desde Leibniz,[94] han observado que los requerimientos formales de las leyes inequívocas y constantes son, en muchos aspectos, más importantes que la necesi-

[93] Needham (1956), pp. 521 *ss.*

[94] *Véase* Elster (1975a), p. 142, para las ideas de Leibniz sobre este punto. Un decidido partidario de esta idea fue Hume; *cf.* por ejemplo, la exposición de sus ideas en Miller (1976), pp. 162 *ss.* Tanto Leibniz como Hume citan la clásica versión de Jenofonte, en que un muchacho de alta estatura, con una corta casaca tiene que cambiar de casaca, por la fuerza, con un muchacho pequeño, que llevaba una casaca demasiado larga para él, y ambos concluyen que en una visión general de la justicia, esto es inaceptable porque socava la institución de la propiedad.

dad de leyes justas, porque si podemos predecir las decisiones del tribunal podemos tomar las medidas precautorias que nos protejan de leyes injustas.

F. Sejersted [95] ha argüido que el imperio de la ley en la Noruega del siglo XIX fue de una solución a los problemas del despotismo precedente. Este imperio de la ley fue amenazado, después, por un nuevo desarrollo democrático. El despotismo y una democracia sumamente desarrollada, para no decir una democracia directa, se basan en el derecho de intervención en todos los dominios, mientras que la democracia constitucional está fundada sobre un conjunto de instituciones estables que no pueden ser anuladas, una vez establecidas. A este respecto, una idea fundamental es la función de la *asamblea constituyente* que establece las reglas básicas que deberán seguir todas las generaciones posteriores. En realidad, sólo la asamblea constituyente es un actor político, en el fuerte sentido de la *politique politisante;* todas las generaciones anteriores están restringidas a la *politique politisée* o la cotidiana puesta en vigor de las reglas fundamentales. La nación puede atarse "a sí misma" (idea controvertida) mediante la asamblea constituyente, confiando ciertos poderes de decisión a la rama judicial, requiriendo que las reglas básicas sólo puedan ser cambiadas por una mayoría de dos tercios o tres cuartos, etcétera.

La *paradoja de la democracia* puede expresarse así: cada generación desea ser libre de atar a su sucesora, sin estar atada por sus predecesoras. Esta contradicción tiene una estructura similar a lo que se ha llamado la contradicción central del capitalismo: cada capitalista desea ba-

[95] Sejersted (1978); también (1973).

jos salarios para los propios trabajadores (lo que causa grandes ganancias) y altos salarios para todos los demás trabajadores (lo que crea una alta demanda).[96] En ambos casos, es posible para cualquier generación (o para cualquier capitalista) comerse su pastel y conservarlo, pero todas las generaciones (o todos los capitalistas) no pueden alcanzar simultáneamente este objetivo. El eslabón con el problema de las preferencias incongruentes de tiempo (*supra*, sección 5 del presente capítulo) es obvio. El equivalente de la estrategia incongruentemente irracional tal vez pueda encontrarse en la reciente política china, especialmente en las fases revolucionarias en que "el plan pierde importancia ya que constantemente lo están revisando".[97] La estrategia de Ulises consiste en comprometer a las generaciones posteriores estableciendo una constitución que incluye cláusulas que impiden cambiarla fácilmente. La asamblea constituyente tiene un carácter excepcional y privilegiado, no por derecho sino por accidente histórico. En situaciones históricas excepcionales e impredecibles, la representatividad de las personas y la legitimidad de los métodos de voto se deciden en el lugar; la drástica ruptura con el pasado deja a la asamblea libre de atar el futuro. Desde luego, la condición importante es que la repercusión debe llegar de fuera del sistema político; queda excluida toda simple operación de "ayudarse a sí mismo".

Es interesante que varios pensadores políticos hayan considerado la posibilidad de una asamblea constituyente periódica, de modo que cada quien,

[96] Robinson (1956), p. 78; *véase* Elster (1978a), cap. V, para otro análisis de la estructura lógica de esta contradicción.

[97] Suttmeier (1974), p. 91.

una vez en su vida, tuviera derecho de voto en los problemas más básicos de la sociedad en lugar de quedar bajo la tutela de generaciones pasadas.[98] Esta intención, aunque loable, parece albergar una contradicción irreductible. Supongamos, por ejemplo, que la asamblea constituyente ha determinado *1)* que habrá asambleas constituyentes a intervalos de 30 años y *2)* que los cambios de la constitución entre estas asambleas deberán contar con una mayoría de dos tercios. Esto nos lleva a preguntar qué mayoría se necesita en las asambleas ulteriores, y quién la decidirá. Si la asamblea constituyente original estableciera el método de voto para asambleas ulteriores, esto significaría que, después de todo *sí* era privilegiada. Y si las propias asambleas ulteriores decidirán la cuestión, inmediatamente caemos en una regresión infinita, pues, ¿cómo se elegirá el método de votación? (Además, queda la cuestión de si una asamblea ulterior podría suprimir el sistema de las asambleas periódicas.) La razón de que esta regresión infinita no necesariamente surja en la asamblea original se encuentra en el carácter excepcional y carismático del grupo, que tiene su legitimidad en circunstancias objetivas fuera de sí misma. Estas razones favorecen la unidad y la unanimidad ante las cuales se evaporan las cuestiones procesales. (Desde luego, esta es una historia idealizada, pero no, creo yo, imaginaria.) La implicación de este análisis es que las generaciones posteriores no tienen la obligación de sentirse atadas por sus predecesoras, pero tampoco tienen ningún derecho legítimo de atar a sus sucesores. La constitución sólo sigue siendo legítima porque todas

[98] Sejersted (1978) cita a Thomas Jefferson y al noruego Hernik Stenbuch.

las alternativas carecen de legitimidad; es un punto focal en el sentido de Schelling o una "línea brillante" en el sentido de Ainslie *(véase* la sección 9 del presente capítulo). Si, digamos, en la asamblea constituyente 70 % votaran para una requerida mayoría de dos tercios para modificar la constitución, mientras que hoy, 60 % está en favor de una mayoría sencilla, ésta no es una base legítima para el cambio.

Examinemos más minuciosamente este problema. En cualquier momento de tiempo podemos suponer que hay un porcentaje bien definido $y = f(x)$ de la asamblea que quiere que la mayoría requerida para un cambio en la constitución sea al menos de x porcentaje. Puede suponerse que y es una función decreciente de x; por simplicidad, también presupondremos continuidad. (Esto puede interpretarse en términos de individuos que aun no se deciden por completo, y que pueden votar por un propuesto cambio de acuerdo con cierta probabilidad, entre 0 y 1.) Si $f(50) > 50$ y $f(100) < 100$, debe haber algún punto fijo $\bar{x}$ entre 50 y 100, de modo tal que $\bar{x}$ porcentaje de la asamblea quiere que la mayoría requerida al menos sea de $\bar{x}$ porcentaje. Podríamos entonces argüir que esta $\bar{x}$ debe ser incorporada a la constitución misma, como el porcentaje más alto en apoyo de ella. Además, podríamos argüir que $\bar{x}$ está actualmente siendo modificada para tomar en cuenta los cambios de las actitudes del pueblo hacia la importancia relativa de la democracia y la estabilidad.

Sin embargo, diríase que tales intentos estaban condenados al fracaso. No tendría ningún objeto atar a las generaciones futuras si pudiesen desatarse ellas solas de esta manera. La única solución, tal vez, sería incorporar una cláusula a la

constitución que no podría ser modificada por ninguna generación, es decir, que si en el año t el porcentaje requerido para hacer cambios a la constitución se modifica, a su vez, de acuerdo con el procedimiento anterior, entonces este cambio no debiera entrar en efecto antes de $t + 30$. Sólo de esta manera podría quedar seguro que el porcentaje $f(x)$ realmente reflejara la actitud del pueblo hacia la democracia, en lugar de ser una racionalización de alguna meta (es decir, un cambio constitucional preciso) que desearan alcanzar *hoy*. Suponiendo que la tendencia a lo largo del tiempo, es hacia valores más bajos de $\bar{x}$, este procedimiento podría dar resultados aceptables. En cambio, si el estado de la opinión vacila entre fundamentalismo y radicalismo en la interpretación del derecho constitucional, el procedimiento tendría entonces la extraña característica de colocar a una generación conservadora en la situación en que hay demasiado espacio para el cambio constitucional; lamentaría no estar más firmemente atada al mástil. Y nuevas reflexiones revelan más paradojas. Una generación liberal desearía imponer unas medidas iliberales para prevenir a la generación iliberal que, según se espera, se seguirá de la aprobación de unas medidas iliberales para atar a la generación liberal, que se espera que siga al siguiente paso, etcétera. Hay una inestabilidad inherente a todas las alternativas a la constitución legada del pasado, que conserva su legitimidad a falta de algo mejor.

El segundo conjunto de problemas que analizaremos bajo el rubro de compromiso previo político concierne a la naturaleza del Estado capitalista según la teoría marxista. Es sabido que durante el auge del capitalismo clásico, el poder

político no estuvo en manos de la clase capitalista. Deberemos interpretar esto como que los capitalistas abdicaron el poder, y, de ser así, ¿en qué sentido y con qué propósitos? Además de Marx, mencionaremos algunos otros autores.

En el nivel más general (como se encuentra, sobre todo en *La ideología alemana)* la teoría marxista del Estado puede plantearse de la siguiente manera. La teoría política clásica a partir de Hobbes, había presupuesto que el Estado era un medio de realizar la solución cooperativa a un Dilema del Preso *jugado por todos contra todos.*[99] Marx rechazó este enfoque como excesivamente general. En lugar de considerar la sociedad como integrada por átomos idénticos con intereses estructuralmente idénticos (aun si sustantivamente opuestos), arguyó que había que descomponer la sociedad en dos o más clases distintas, cada cual con su interno Dilema del preso. Arguyó que, además, el Estado capitalista tiene la doble tarea de facilitar la solución cooperativa al Dilema del preso al que juegan los capitalistas unos contra otros, y de prevenir que la solución cooperativa (organización y solidaridad) surgieran en el Dilema del preso, jugado por los trabajadores unos contra otros.[100] (Además, pue-

[99] Para esta perspectiva, *véase* Baumol (1952), Olson (1965) y M. Taylor (1976a).

[100] En términos correctos, la tarea del Estado es facilitar los carteles y organizaciones de patronos y prevenir la formación de sindicatos. En apariencia, las English Combination Acts fueron dirigidas contra patronos y trabajadores por igual, pero en primer lugar hubo una asimetría formal en el hecho de que los estatutos "sólo permitían una acción civil contra el patrono que violara el contrato pero por lo contrario permitían acción penal contra los obreros que violaran el contrato" (Marx 1867, p. 740), y en segundo lugar, las Acts eran aplicadas en

de haber algunas tareas que van en interés de ambas clases, como la provisión de bienes públicos, pero Marx arguyó que el desarrollo del capitalismo haría que estos fuesen asumidos cada vez más por la empresa privada.)[101] En este contexto, no nos interesa el aparato del Estado para reprimir a los trabajadores. (Sólo nos interesaría si fuese establecido por los propios trabajadores, como se sugiere en la última cita de Kydland y de Prescott, *supra.*) Por consiguiente, limitaremos nuestra atención al Estado capitalista como arma para proteger el interés de la clase capitalista en general contra los intereses de sus miembros individuales y para proteger los intereses a largo plazo de la clase contra los intereses a corto plazo (no trataré aquí de los interesantes problemas que surgen cuando los intereses individuales a corto plazo y los intereses colectivos a corto plazo consiguen, oponiéndose ambos al interés colectivo a corto plazo, como en el caso del "imperialismo del libre comercio").[102]

Un ejemplo en que el interés de clase entró en conflicto con el interés individual nos lo ofre-

grados muy distintos contra patronos y trabajadores que violaran los contratos.

[101] Marx (1857-1858), pp. 429-430. Marx no especifica el mecanismo que pudiera crear esto, pero al menos se nos ocurren dos posibilidades. O bien algunas de las empresas podían alcanzar unas dimensiones que harían lucrativo para ellas aportar los artículos por su cuenta (Olson 1965, p. 29) o podían inventarse recursos institucionales (como el sistema de patentes) para internalizar las economías externas. Marx ciertamente exageró esta tendencia, pero hay una verdad importante en la idea de un capitalismo que constantemente *reprivatiza* las actividades que fueron emprendidas por el Estado, porque el sistema de mercado no las aportó.

[102] *Cf.* Elster (1978a), cap. v, para un análisis de esta cuestión.

ce la legislación fabril inglesa, que, lejos de ser un freno a la voracidad capitalista, fueron

> la expresión negativa de la misma voracidad. Estas leyes contienen la pasión del capital por una ilimitada sangría de la fuerza de trabajo, limitando obligatoriamente la jornada laboral por regulaciones estatales, hecho por un Estado que es gobernado por el capitalista y el terrateniente. Aparte del movimiento de la clase obrera que día tras día se hizo más amenazador, la limitación del trabajo fabril fue dictada por la misma necesidad que cubre de guano los campos de Inglaterra. La misma ciega avidez de saqueo que en un caso agotó la tierra, en el otro había desarraigado la fuerza misma de la nación.[103]

Tales casos son relativamente fáciles de comprender. Como en la cartelización forzosa[104] o en las leyes de expropiación,[105] la burguesía puede com-

[103] Marx (1867), pp. 238-239.

[104] *Véase* Mangset (1974k) para un relato de este fenómeno en Noruega, entre las dos guerras.

[105] Marx (1845-1846), p. 340. Contra Stirner, quien veía la existencia de estas leyes como prueba de que el Estado era el verdadero propietario de la propiedad "privada", Marx opone una serie de notables formulaciones. Stirner convierte las *contradicciones* de la propiedad privada en la *negación* de la propiedad privada. Su argumento equivale a decir que mi perro guardián es el verdadero propietario de mi casa. Y por último, en una frase en que la doble tarea del Estado (proteger a la clase capitalista contra la clase obrera y contra sus propios miembros individuales), queda claramente expresada: "Weil die Bourgeois dem Staat nicht erlauben, sich in ihre Privatinteressen einzumischen, und ihm nur soviel Macht geben, als zu ihrer eignen Sicherheit und der Aufrechthaltung der Konkurrenz nötig ist, weil die Bourgeois überhaupt nur insofern als Staatsbürger auftreten, als ihre Privatverhältnisse dies gebieten, glaubt Jacques le bbonhomme, dass sie vor dem Staate 'Nichts sind'" (Marx, 1845-1846, p. 339).

prender que va en su propio interés el que se le impida buscar su propio interés de manera irrestricta, y atarse a sí misma por consecuencia. Otros casos mencionados de abdicación capitalista son más complejos y ambiguos. En el caso que acabamos de citar, estábamos enfrentándonos a un Estado gobernado por *capitalistas* (aunque Marx añade "y terratenientes") que aplica medidas en bien de la clase capitalista en conjunto. En otros textos más importantes, Marx reconoció que el Estado *no* es un Estado capitalista en el sentido directo de ser gobernado por capitalistas para sus propios intereses. Por lo contrario, el poder político estaba firmemente en manos de la aristocracia, como en Inglaterra, o de un régimen cesarista-burocrático, como en Francia. Esta separación del poder económico y del político obviamente crea un problema para una teoría que tiene como afirmación básica que "el ejecutivo del Estado moderno no es más que un comité para administrar los asuntos comunes de la burguesía".[106] Va en crédito de Marx el que suprimiera este "más que"-ismo cuando procedió al análisis histórico detallado. Sin embargo, queda por demostrar que el Estado moderno —además de ser muchas otras cosas— *también* sea un comité para administrar los asuntos comunes de la burguesía.

Marx adoptó dos diferentes líneas de argumento para demostrar esto, para los casos inglés y francés, respectivamente. En los artículos escritos para la *New York Daily Tribune*, la *Neue Oder Zeitung* y la *Presse* de Viena sobre la política inglesa, Marx repitió una y otra vez que los capitalistas industriales deliberadamente se abstenían de tomar el poder político porque temían que la

[106] Marx y Engels (1848).

concentración de poderes económico y político en una sola clase agudizara la lucha de clases y condujera a una revolución social. Las pérdidas económicas a corto plazo causadas por dejar el poder en manos de la aristocracia parecieron más pequeñas que las ganancias políticas (y las ganancias económicas a largo plazo) logradas por distraer la atención y la revuelta lejos de la clase capitalista, y borrar los lineamientos del conflicto de clases. Marx cita a Juvenal para expresar el dilema de la clase capitalista, desgarrada entre su viejo enemigo y su enemigo naciente: *Et propter vitam vivendi perdere causas.*[107]

Sin embargo, esta explicación difícilmente pasaría por análisis del compromiso previo, porque muchas de las normas dadas en la sección 2 del presente capítulo no se satisfacen. Podemos mostrar esto planteando una confrontación con algunas otras opiniones de la política inglesa y con la propia versión de la política francesa, por Marx. En el *Economist* apareció en 1862 un artículo con el notable título de "La ventaja para un país comercial, de un gobierno no comercial",[108] en que el autor (¿el propio Walter Bagehot?) argüía convincentemente que los intereses comerciales se abstendrían de tomar el poder político si conocieran su propio bien, porque la aristocracia era mucho más capaz que "el gobier-

[107] *Neue Oder Zeitung*, 12 de junio de 1855. Hasta donde yo sé, se ha prestado poca atención a los escritos de Marx sobre la política inglesa, aunque tratan del país capitalista por excelencia y son mucho más congruentes teóricamente que con sus opiniones económicas, que sus escritos sobre la historia política de Francia. La mayor parte de los escritos (pero no todos ellos) se encuentran reunidos en Marx y Engels (1971).

[108] Debo esta referencia, así como otras de este párrafo, a Grindheim (1975).

no de los ricos"[109] de adoptar la visión desinteresada a largo plazo. Similares opiniones fueron expresadas por Hyppolite Taine.[110] El mismo análisis fue propuesto por Schumpeter, quien arguyó que el gobierno de la aristocracia era benéfico para la burguesía porque "necesitaba un amo"[111] para protegerse a sí misma contra sí misma. Estos argumentos es por lo menos discutible que se hayan satisfecho las normas (i), (ii) de la sección 2 del presente capítulo. Todos ellos presuponen que la burguesía deliberadamente limitó o al menos se contuvo de extender su poder político porque no podía confiar en sí misma para mantener el medio estable necesario para el desarrollo a largo plazo. Esta estrategia excluyó algunas opciones que de otra manera no habrían sido factibles, y esta *exclusión fue la razón de ser de la estrategia.* Esta última afirmación no es válida en el caso de Marx; para él, la reducida libertad de acción fue un costo asociado a la política, y no el objetivo que había que alcanzar. Esto significa que no se satisfizo la norma (i). La burguesía se abstiene del poder para manipular a la clase obrera y no para controlarse a sí misma por medio de un amo autoimpuesto.

Parece perfectamente claro que ni la visión que Marx tuvo de la política inglesa ni las otras teorías mencionadas satisfacen la norma (v) de la sección 2 del presente capítulo. Absurdo sería decir que los capitalistas industriales abdicaron del poder; nunca lo habían tenido. No se puede argüir siquiera con mucha verosimilitud que se tomara una decisión deliberada de no tomar el

[109] Bagehot (1966), p. 122.
[110] Citado por Grindheim (1975).
[111] Schumpeter (1954), p. 139.

poder, aun cuando el hecho de que la Liga contra la Ley de Granos no se convirtiera ella misma en un partido político podría interpretarse como una decisión de esta índole.[112] Antes bien, la mejor conceptualización sería decir que la aristocracia quedó libre de maximizar sus propios objetivos con una ganancia razonable para la clase capitalista como limitación indirecta. Esto implica, sin duda, que la clase capitalista habría tomado el poder si sus intereses hubiesen estado en grave peligro, pero esta fórmula contrafáctica no ofrece la clara prueba necesaria para satisfacer la norma (v), *supra.*

Ahora bien, Marx nunca pretendió que su versión fuese un análisis del compromiso previo, por lo que no es posible criticarlo por no haber satisfecho las normas que entran en esta noción. Antes bien, su error consistió en escoger un actor que no era, o utilizar el concepto erróneo de un actor. En el sentido convencional del término, sólo la aristocracia, con sus fracciones y conflictos internos, puede considerarse como un actor político en la política inglesa de la época. Los capitalistas y los obreros ciertamente ejercían una influencia sobre la política, pero más como limitación que búsqueda de un objetivo. Ciertamente Marx consideró que la clase capitalista buscaba su objetivo económico por medios políticos, y la no participación en la política fue interpretada entonces, intencional e instrumentalmente, como un retiro de la política. Aquí Marx, lejos de descuidar la perspectiva del actor en la política, tendió a exagerar los aspectos racionales de la conducta capitalista.[113] Lo mismo

[112] Debo esta sugerencia a Kare Tønnesson. Marx establece el mismo punto en su artículo sobre los cartistas, *New York Daily Tribune,* 25 de agosto de 1852.

puede decirse de los otros autores antes citados. Schumpeter tal vez tenga razón cuando dice que fue racional (es decir, ventajoso) para la burguesía mantenerse apartada del poder, pero no podemos concluir que por ello se mantuvo apartada del poder.

Desde luego, queda la cuestión de la libertad de decisión de la aristocracia dentro de las limitaciones definidas por las otras clases. Si una clase logra imponer limitaciones lo bastante fuertes a los que toman las decisiones, entonces esa clase puede ser el actor político central en un sentido menos convencional del término. Por ejemplo, en este sentido, podemos decir que los obreros polacos de hoy son actores políticos, aun si no están organizados en el sentido tradicional.[114] Lo mismo puede decirse de muchos motines y amotinados a lo largo de la historia.[115] Estamos tan acostumbrados a la dictomía hegeliano-marxista entre clases que existen *en sí mismas* y clases que existen *por sí mismas* que olvidamos la importancia de la etapa intermedia en que las clases ejercen su influencia por medio de su existencia *para otras*.[116] Si mediante esta existencia

[113] *Cf.* en particular su conspiratoria interpretación de la política de Palmerston, en Marx (1971), pp. 278-279.

[114] "La verdadera amenaza para el gobierno no es la Iglesia ni los intelectuales sino los trabajadores. Nadie sabe cuál será probablemente su reacción en ninguna situación. Un destacado sociólogo habla de ellos como 'las masas políticamente desorganizadas' y dice que su existencia misma limita la libertad de movimiento del gobierno" (*Sunday Times*, 25 de septiembre de 1977).

[115] Hobsbawm (1959), cap. VII.

[116] *Cf.* Evans-Pritchard (1940, p. 120): "Un hombre de una tribu ve al pueblo de otra tribu como un grupo diferenciado para el cual tiene una pauta indiferenciada de conducta mientras se ve a sí mismo como miembro de un segmento de su propia tribu."

refleja logran algo parecido a lo que habrían podido tener por medio de organización, podemos referirnos a ellas como "casi-actores". Ocasionalmente un grupo puede alcanzar más como casi-actores que como actores y tratar de atarse a sí mismo renunciando a la condición de actor. *Qua* participantes responsables en el toma y daca político podrían obtener menos que por medio de su valor peligroso *qua* potenciales amotinados. Esta idea se relaciona con la observación de T. S. Schelling de que un negociador estúpido puede obtener más concesiones que uno sagaz que sea capaz de ver la justicia como pretensión del adversario.[117]

Después de esta digresión sobre la noción de un actor político, me volveré al análisis marxista de la política francesa, que ciertamente pretendió *ser* un análisis del compromiso previo. La explicación marxista del golpe de estado de Louis Bonaparte tiene mucho en común con la visión de la política que empieza con Schumpeter, con el rasgo añadido de que como acto de abdicación, puede demostrarse empíricamente que ocurrió. La burguesía francesa necesitaba un amo; lo encontró en Louis Bonaparte; abdicó del poder transfiriéndoselo a él.

> Por tanto, cuando la burguesía desprecia como "socialista" lo que antes ensalzaba como "liberal", confiesa que su propio interés le ordena esquivar el peligro de su "gobierno propio" que para poder imponer la tranquilidad en el país tiene que imponérsela ante todo a su parlamento burgués, que para mantener intacto su poder social tiene que quebrantar su poder político; de los individuos burgueses sólo pueden seguir explotando a otras clases y disfrutando apaciblemente de la propiedad, la fa-

[117] Schelling (1963), pp. 22 *ss.*

> milia, la religión y el orden bajo la condición de que su clase sea condenada con las otras clases a la misma nulidad política; que para salvar la bolsa, hay que renunciar a la corona, y que la espada que había de protegerla tiene que pender al mismo tiempo sobre su propia cabeza como la espada de Damocles.[118]

Sin duda, es un simple accidente el que este magnífico pasaje retórico no añada la metáfora de Ulises y las sirenas a los demás; sin embargo, *es* retórico, y no es un análisis. Aun si concedemos a Marx el punto de que el carácter anticapitalista del régimen bonapartista fue benéfico o hasta indispensable para la supervivencia del capitalismo como sistema económico, aun queda por demostrar que esto *explica* el surgimiento y mantenimiento de tal régimen. Hablando en abstracto, dos de tales explicaciones son posibles, de las variedades funcionalista y la intencional. Me remito al capítulo I, sección 5 *supra*, para la sugestión de que el carácter anticapitalista del Estado bonapartista puede explicarse por su función para el modo de producción capitalista. En cuanto al modo intencional de análisis, requeriría prueba empírica de que el régimen bonapartista deliberadamente se estableció por un acto de abdicación del poder por la clase capitalista. Hasta donde yo sé, no contamos con tal prueba. La razón por la que parecería que si la norma (v) de la sección 1 del presente capítulo se cumple en este caso es la indiscutible discontinuidad del *golpe de Estado*. La continuada supremacía política de la aristocracia inglesa podría ser considerada, cuando mucho, como un acto de omisión de la burguesía inglesa, pero el súbito cambio

[118] Marx (1852), sec. IV.

del poder, de un régimen liberal a uno bonapartista satisface al menos una condición para que éste sea un acto de comisión por la burguesía francesa. Sin embargo, al menos es igualmente verosímil considerar el verdadero actor en el verdadero Louis Bonaparte. Decir que la debilidad de la burguesía permitió su ascenso al poder no es decir que la burguesía le permitió tomar el poder. Abdicar inmediatamente antes de ser destronado puede mejorar las apariencias, pero no modifica la sustancia.

Concluyo, entonces, que el análisis de la democracia ha ofrecido algunos ejemplos convincentes de compromiso previo político; no así el análisis del Estado capitalista. Debemos subrayar la idea fundamental de la asamblea constituyente, no como recurso ficticio, como en las teorías del contrato del Estado, sino como una verdadera asamblea histórica que trataba de atar a sus sucesoras. Esta es la más cercana analogía en sociedad al estado de ánimo de Ulises en esa dramática parte de su jornada.

9. Algunas conclusiones y nuevas preguntas

Revisemos la gama de estrategias de compromiso previo que hemos analizado en el curso de este capítulo:

1) Manipulación del conjunto factible
  a) Restricción del conjunto de acciones físicamente posibles
  b) Cambio de la estructura de recompensas por apuestas públicas indirectas

2) Manipulación del carácter
  a) Fortalecimiento de la fuerza de voluntad
  b) Cambio de la estructura de preferencia

3) Manipulación de la información
   a) Cambio del sistema de creencias
   b) Evitar la exposición a ciertas señales

A esto debemos agregar una estrategia que no satisfaga la norma (iii) de la sección 2 del presente capítulo, pero que también es una respuesta al problema de la flaqueza de voluntad:

4) Manipulación por una redisposición del espacio interior
   a) Utilizando apuestas privadas indirectas
   b) Utilizando planificación congruente

Comentemos brevemente estas estrategias y subestrategias. Dentro del conjunto de acciones que cambian el conjunto factible yo distingo entre, por ejemplo, la estrategia de irse a dar un largo paseo en las montañas para no poder contar, físicamente, con cigarrillos, y la estrategia de decir a nuestros amigos que dejaremos de fumar, para modificar el sistema de recompensas. Esta última estrategia también produce un cambio en el conjunto factible, porque la opción "continuar fumando sin comentarios sarcásticos" queda ahora fuera de nuestro alcance. Dentro de las acciones que modifican el carácter yo distingo entre la estrategia general de fortalecer la fuerza de voluntad y la estrategia más específica de modificar algún deseo particular: la primera nos permite ascender gradientes más altas mientras que la segunda reduce la altura de la gradiente que habremos de ascender. Dentro del conjunto de acciones que modifican la información sobre la cual se toman nuevas decisiones, yo distingo entre la estrategia radical de producir nuevas creencias fácticas (incluyendo una estrategia para causar olvido acerca de la inducción) y la estrategia mo-

derada de protegernos de ciertas señales o claves que haya en el medio. *No* se trata de evitar la exposición a información en el sentido estricto, que es tan paradójica como la inducción de nuevas creencias;[119] sólo es cuestión de evitar la exposición a condiciones que pudiesen desencadenar la eficacia de una información que ya estaba en nuestra profesión. No podemos evitar deliberadamente las tabaquerías con el propósito de producir olvido acerca del tabaco. Por lo contrario, la necesidad de estar constantemente en estado de alerta ante tales tiendas para cruzar la calle antes de llegar a ellas tenderá, más probablemente, a fortalecer nuestra conciencia del tabaco. No obstante, la disposición al tabaco en concreto puede ser más tentadora que la conciencia del tabaco en abstracto, de modo que podemos aceptar más de esto último, como costo de tener menos de lo primero.

Las más importantes de estas estrategias son probablemente los métodos (1b) y (2b). En muchos casos reales, la estrategia (1a) es infactible o entraña sacrificios excesivos y por tanto no satisfaría la norma (1b) de la sección 2 del presente capítulo. Lo mismo, creo yo, puede decirse del método (2a), que a menudo entrañaría en forma de "exceso de matanza" cuya factibilidad acabaría con el problema mismo. Sin embargo, debemos notar la profunda sugestión de Ainslie de que las estrategias (4a) y (2a) pueden estar íntimamente relacionadas, es decir, que la técnica de las apuestas privadas indirectas puede implicar un general reforzamiento de la fuerza de voluntad. Cuando el sujeto redispone su espacio interno y sus sistemas de recompensas, de

[119] Para las paradojas de evitar información, *véase infra*, cap. IV, sec. 4.

modo que el no actuar de acuerdo con una decisión (es decir, la decisión de dejar de fumar) tiene implicaciones negativas para otras decisiones (por ejemplo, la decisión de hacer dieta), el sistema interrelacionado de objetivos y deseos puede llegar a mostrar la rigidez y la inflexibilidad tan a menudo asociadas a la idea de fuerza de voluntad.[120] En cuanto a la estrategia (3a), parece tan costosa que sólo se le puede escoger en unos cuantos casos extremos; la estrategia (3b) es utilizada con frecuencia, pero es de dudosa eficacia.[121]

Yo supongo que en casi todos los casos cotidianos, la estrategia más eficiente incluye una combinación de los métodos (1b) y 2b). Comprendiendo que por medio de una serie de acciones puedo alcanzar un *hexis* a partir del cual se darán naturalmente las condiciones deseadas ("sans violence, sans art, sans argument") y que cada una de estas acciones está fuera del alcance inmediato de mi fuerza de voluntad, puedo comprometerme previamente con ellas, cambiando el sistema de recompensas. Así es como muchas personas logran dejar de fumar. Producir el estado de no adicción simplemente absteniéndose de fumar puede ser demasiado difícil; establecer una permanente de apuestas públicas indirectas es complicado, y en todo caso es innecesario si se quiere que un sistema temporal produzca este estado y, por ello mismo, se haga superfluo.

[120] Ainslie (1977) analiza esto extensamente.

[121] Ainslie (1975), p. 478. Una evaluación más positiva se encuentra en la introducción de los compiladores a Mahoney y Thoresen (comps.) (1975), pp. 40 *ss.* Sin embargo, con base en su versión, es claro que el "control de estímulo" llega a cubrir otras estrategias, como cuando se refieren a la regla "siempre ve a comprar alimentos después de una comida abundante".

"Después de los 40 años, cada quien es responsable de su cara." Esta verdad, igualmente difícil de confrontar que de disputar, no es más que un aspecto del hecho general de la *responsabilidad de sí mismo.* Nadie puede decir jamás "Así es simplemente como soy" como excusa para su conducta, pues cada quien habría podido convertirse en un tipo diferente de persona. En la bibliografía filosófica, esta responsabilidad ha sido afirmada desde dos distintos puntos de vista. Uno de ellos, de origen inicialmente aristotélico, coloca al ego abiertamente en el mundo de los procesos causales y arguye que puede y debe forjarse a sí mismo explotando estos procesos. Este es el enfoque explorado en este capítulo. Otro, representado por Sartre y más recientemente por Charles Taylor, presupone que es posible cambiarse a sí mismo redefiniendo y redescribiendo el ego. Esto parecería más cerca de las técnicas de las apuestas privadas indirectas y de la planeación congruente. Como tantos otros escritores, Taylor ilustra su enfoque con el problema de la dieta:

> Tomemos el caso del hombre que está combatiendo la obesidad y al que se convence de que la vea como cuestión simplemente cuantitativamente más satisfactoria y no como cuestión de dignidad y degradación. Como resultado de este cambio, su lucha interna misma se transforma, ahora es una experiencia totalmente distinta. Las motivaciones opuestas —el amor al pastel de crema y su insatisfacción consigo mismo por tal indulgencia— que son los "objetos" que están pasando aquí por una redescripción, no son independientes... Cuando llega a aceptar la nueva interpretación de su deseo de controlarse a sí mismo, este deseo mismo se ha alterado. Cierto, puede decirse en determinado nivel que tiene la misma meta,

> que deja de comer pastel de crema pero como ya no se interpreta esto como busca de dignidad y respeto propio, se ha convertido en un tipo totalmente distinto de motivación...
>
> Así, nuestras descripciones de nuestras motivaciones, y nuestros intentos de formular lo que nos parece importante, no sólo son descripciones ya que sus objetos no son plenamente independientes. Y sin embargo, tampoco son simplemente arbitrarias, de modo que todo se vale. Hay interpretaciones más o menos adecuadas, más o menos ciertas, más clarividentes o más autoengañosas.[122]

Dentro de ciertas limitaciones, esto dice el dicho: esta es la verdad importante expresada aquí. En las ciencias sociales esto se ha estudiado bajo el marbete de predicciones que se cumplen a sí mismas, y el teorema del punto de fijo de Brouwer se ha invocado para mostrar que al menos en algunos casos importantes es posible describir la situación de un modo que tome en cuenta el hecho de que la descripción modifica el estado que se está describiendo.[123] El punto central del pasaje de Taylor, al que nada corresponde en estos estudios, es que puede haber bastantes puntos fijos, varios modos mutuamente incompatibles de describir la situación cada uno de los cuales se vuelve cierto por virtud de ser afirmado. Desde luego, hay limitaciones que impiden que cualquier descripción se realice por sí mis-

[122] C. Taylor (1976), p. 295; también C. Taylor (1971) establece el mismo punto.

[123] *Cf.* n. 25 *supra.* Para un análisis reciente con referencias completas a la bibliografía económica y sociológica, *véase* Brams (1976), cap. III. Hasta donde yo sé no se ha hecho ningún intento por vincular estas cuestiones con los problemas análogos que surgen en psicología, aunque la noción de *equilibrio reflexivo,* elaborada por John Rawls (1971) podría interpretarse en términos de un razonamiento de punto fijo.

ma. Estas corresponden, supuestamente, a rasgos de carácter de orden superior y como tales deben ser sujeto de la tesis de que el hacerlo lo hace, dentro de otras limitaciones, etcétera. Decir "Así es como soy" es en parte descripción, en parte descubrimiento y en parte creación.

Los enfoques aristotélico y sartreano se complementan uno al otro en lugar de contradecirse, al menos si estamos dispuestos a atribuir más importancia a las limitaciones que el propio Sartre. La estrategia del compromiso previo puede utilizarse para realizar un estado que no es alcanzable mediante simple redefinición. Yo, por mi parte, creo que los límites puestos a las descripciones que se realizan por sí mismas son muy fuertes, y se vuelven cada vez más cuando pasamos de la descripción del primer orden de los rasgos de carácter a las descripciones de orden superior de las limitaciones de primer orden. Sin embargo, cualquiera, reconociendo que *algunos* límites del principio del decir lo hace, deben dar *cierto* espacio a los recursos para atarse a sí mismo. La figura que puede verse como un pato y como un conejo, pero no como ambos simultáneamente, puede servirnos aquí de analogía.[124] Las líneas de la figura sirven como limitación a la forma en que puede verse la figura, pero añadir una línea extra puede cambiar esta imitación de una manera que puede ser práctica. Aquí, el "acto" de pasar de la visión pato a la visión conejo corresponde a una redefinición de la situación, y el hecho de añadir una línea extra a la estrategia del compromiso previo. La responsabilidad del ego incluye ambos tipos de acción. (En el cap. IV, sec. 3 la redefinición del ego median-

[124] Debo esta interpretación de Sartre a Dafginn Føllesdal.

te algún otro redefinido sirve como uno de los puntos de partida para un análisis del amor y en el cap. IV, sec. 4 las mismas nociones afectan el análisis del autoengaño.)

En conclusión, deseo mencionar dos problemas perturbadores a los que no les veo solución fácil. El primero puede resumirse en una frase de William James, en el sentido de que "la vida ética *más elevada*... consiste en todo momento en la violación de reglas que se han vuelto demasiado estrechas para el caso real".[125] La forma más baja de la vida ética es, supuestamente, la impulsividad total y miope. A un nivel superior es la vida de acuerdo con reglas o estrategias autoimpuestas de compromiso previo; y al más alto nivel, la deliberada violación de estas reglas, cuando había que considerar todas las cosas, esto parece justificado. Para ver más claramente dónde está el problema en este análisis, podemos distinguir entre cinco sentidos de "acción no gobernada por reglas":

1) Acciones efectuadas por una persona que carece del concepto de regla
2) Acciones desempeñadas por una persona que típicamente actúa de acuerdo con las reglas, pero que, o bien
  a) no sigue la regla en un caso particular o
  b) decide violar la regla en un caso particular

3) Acciones efectuadas por una persona que ha decidido no actuar nunca de acuerdo con reglas y que, o bien
  a) Nunca ha seguido las reglas o bien
  b) ha seguido las reglas y decidido abandonarlas por completo.

[125] Citado en Ainslie (1977), p. 38, quien también se refiere al requerimiento de "tolerancia de la ambigüedad" como la marca de una persona adulta.

Para ver la pertinencia de estas distinciones, considérese el problema del verso libre contra el verso rimado. La mayoría de los poetas que escriben versos rimados argüirían, creo yo, que esta es una limitación autoimpuesta y necesaria, y que la ilimitada libertad del verso libre es más destructiva que creadora. Tan sólo limitando el conjunto infinito de afirmaciones posibles al subconjunto más tratable (pero aún muy grande) de afirmaciones que satisfagan los requerimientos formales de la rima y del metro puede el poeta crear el medio en que ejercer su don de elegir entre tales afirmaciones. La forma estricta del soneto permite al poeta "poner el Caos en 14 versos y mantenerlo allí";[126] y una idea similar subyace, sin duda, tras la hazaña recién efectuada por George Perec al describir toda una novela en que no se emplea ni una sola vez la letra "e".

Contra esta línea clásica son posibles muchas otras actitudes. Algunos dirían que para escribir verso libre hay que haber dominado la técnica del verso rimado (caso 3b), y recibirían con escepticismo al tipo de poeta que se atribuyera la capacidad de prescindir de esta maestría (caso 3a). Sin embargo, escritores de ambas persuasiones convendrían en que la poesía es esencialmente distinta de la simple charla (caso 1) y que las reglas —si tiene que haberlas— deben ser dominadas por completo (en contraste con el caso [2a], correspondiente a la flaqueza de voluntad). La posición más compleja sería que la más alta vida poética se alcanza mediante el verso rimado, roto por una ocasional regularidad cuando se requiera. *Sí* es importante atarse a sí mismo, escapar del abismo de las posibilidades ilimita-

[126] Edna St. Vincent Millay (1975), soneto clxviii "Put Chaos into fourteen lines and keep him there".

das, pero no menos importante es tener olfato para las ocasiones cuando la perfección exige una libertad de movimiento sin el peso de las reglas, como en el caso (2b).[127] Esta última actitud parece también captar la idea expresada por William James, aunque el caso (3b) también puede ser una interpretación posible.

En el caso ético, en contraste con el estético, el principal problema consiste en distinguir entre los casos (2a) y (2b); cómo distinguir entre violar las reglas por buenas razones y violarlas por malas razones. ¿Cómo podemos saber —y cómo pueden otros saber— si un ejemplo dado de comportamiento impulsivo (no gobernado por reglas) corresponde al nivel más alto o más bajo de la mente? Las excepciones reconocidas de antemano como excepciones no presentan dificultades. Lo que queremos es una norma para lo que *habría sido* reconocido de antemano como excepción legítima si se hubiese planteado la cuestión. Ainslie sugiere[128] que en tales casos necesitamos una *línea brillante* (o punto focal) en nuestro espacio interno para decirnos si una excepción determinada es una racionalización *ad hoc* o bien es auténtica. Aquí, puede ser importante la internalización de las normas parentales, para lograr este ideal de *control sin rigidez*.

El problema también puede verse desde el punto de vista de la persona que tiene la tarea de atarnos: ¿en qué condiciones deberá liberarnos del mástil cuando le roguemos que lo haga? Derek Parfit ha elaborado el siguiente ejemplo:

[127] "¿Hay mayor alegría para un escritor rigurosamente preparado en la disciplina artística, o para un protestante 'limítrofe' como Gide o Eliot, que cobrar reverente conciencia de unas reglas estrictas que algún día podrá deleitarse en violar?" (Peyre, 1967, p. 227).

[128] Ainslie (1977), p. 23 y *passim*.

> Imaginemos un ruso del siglo XIX que, en varios años, heredará vastas posesiones. Como tiene ideales socialistas, se propone hoy ceder la tierra a los campesinos. Pero sabe que con el tiempo sus ideales pueden marchitarse. Para precaverse contra esta oportunidad, hace dos cosas. Primero, firma un documento legal que automáticamente cederá la tierra y que sólo podrá ser revocada con el consentimiento de su esposa. Dice entonces a su esposa: "Si alguna vez cambio de opinión y te pido cambies el documento prométeme que no consentirás." Podría añadir: "Considero que mis ideales son esenciales para mí. Si yo pierdo estos ideales, deseo que tú pienses que *yo* he dejado de existir. Te pido considerar a tu esposo, entonces no como yo, el hombre que te pide esta promesa, sino sólo como su ego posterior. Prométeme que harás lo que él te pida." [129]

Parfit pasa entonces a argüir que en un modo de pensar verosímil: "ella jamás podrá quedar liberada de su compromiso. Pues el ego al que ella se comprometió al tratar de liberarla, dejaría de existir".[130] Esta ingeniosa variación del tema de *Catch 22* (del cual hablaremos más en el cap. III, sec. 9), tal vez no tenga un poder del todo convincente. Plantea las mismas cuestiones éticas que la obligación de cumplir las promesas hechas al lado de un lecho de muerte, con algunas complicaciones añadidas. Sin embargo, estas son cuestiones difíciles y no realmente afines a nuestro problema. En lugar de construir un caso de varios egos sucesivos, como en el ejemplo de Parfit, debemos imaginar una situación en que varios egos coexisten simultánea y jerárquicamente, representando la forma inferior, intermedia y superior de la vida ética como la definimos

[129] Parfit (1973), p. 145.
[130] *Ibid.*, p. 146.

antes. Entonces, la cuestión es saber cómo otras personas pueden escoger un portavoz organizado para la jerarquía; cómo pueden decidir si la revocación de una orden es emitida por el ego superior o el inferior. Otras personas pueden no tener acceso directo a *nuestras* líneas brillantes. Parfit parece creer que nunca debemos ceder, por ejemplo, a una persona que me ruegue darle los cigarrillos que antes me había rogado retirarle,[131] pero esto es cometer petición del principio planteado por Ainslie y James, a saber, que a veces yo puedo tener buenas razones para pedir que me liberen, en el sentido de razones que yo habría aceptado antes de dejarme atar.

El segundo problema es aún más confuso. Se refiere a la gama de regresión infinita que tiene que surgir en cuanto introducimos la gama de la automanipulación. Hemos analizado varios casos de egos de tres filas, y en principio no parece haber razón para que esto no pudiera extenderse a cualquier número de niveles. Cierto que el número tres surge muy a menudo, y que es difícil construir casos de cuatro niveles que resulten verosímiles. (La trinidad freudiana del ello, el yo y el superyó correspondía más o menos a las formas inferior, superior e intermedia de la vida ética, respectivamente.)[132] No obstante, la posibilidad lógica de una jerarquía con infinitos niveles resulta perturbadora, y resulta difícil saber qué contestar a una persona que pida no ser liberada en los términos siguientes: "El ego gobernado por reglas que hoy le habla a usted está a un nivel superior, y no un nivel por debajo de ese ego espléndidamente arrogante que, invocando la idea jamesiana de la vida ética superior, le rogó

[131] *Ibid.*, p. 145.
[132] Ainslie (1977), p. 38.

a usted liberarlo. Le aseguro que esta falta de respeto a las reglas se justifica en el caso de mi superyó neurótico y rígido, pero no en el caso de las reglas morales que yo adopto por motivo del imperativo categórico y a la luz de la razón pura." Y —pasando del punto de vista del observador al del actor—, ¿puedo *realmente* estar seguro de que la idea de James no es una tentación que yo debiera resistir? ¿Cómo puedo saber que mi línea brillante no está desplazándose constantemente "a mis espaldas", para justificar, *ex post facto*, cualquier excepción a la regla? [133]

[133] Para un ejemplo en que muchos lectores se reconocerán a sí mismos, cito de Arthur Stinchcombe (correspondencia privada): "Un problema a lo largo del mismo lineamiento sobre el que yo he reflexionado es el problema de juzgar la productividad del trabajo de los jóvenes investigadores. Supongamos que diferentes personas tienen una esperada productividad de mayores innovaciones, de modo que un distinguido científico puede esperar una innovación importante cada cinco años. Pero en esa circunstancia, los distinguidos científicos fácilmente pueden trabajar diez años con algo rutinario, mientras que uno con una esperada innovación importante cada veinte años puede tener la suerte de hacer la suya en los cinco primeros años. La pregunta es saber si es posible establecer el sistema de tal modo que se reconozcan o predigan nuestros errores. Por ejemplo, la sala común de Oxford puede ser una norma predictiva mejor que las verdaderas realizaciones. Yo recuerdo que Whitehead recomendó a Russell para una beca pasando por encima de un candidato con mejores calificaciones, en un examen. Aquí, la idea es que una norma para mantenerse siendo universalista a veces puede aumentar en lugar de reducir el riesgo de error. Yo no creo realmente que éste sea el caso, habiendo cometido mis errores lo bastante a menudo por ambos motivos. El argumento me recuerda al que Claude Bernard discutió tan eficazmente, de la 'intuición clínica'. Tal vez si alguien pudiera llenar universidades enteras con Whiteheads, y sin embargo impedir que fueran corrompidas por sus poderes y vanidad; pero en este mundo yo dudo de que esto pudiera

No tengo ninguna idea que contara como solución a este problema, pero sí tengo cierta corazonada acerca de las líneas a lo largo de las cuales más probablemente surgiría una solución. Invocaríamos, creo yo, una analogía tomada de la teoría de los juegos, donde la regresión aparentemente infinita de que "Yo creo que él cree que yo creo..." no impide que la teoría tenga una solución única y predecible, aun a falta de estrategias dominantes.[134] Este enfoque es propuesto, por ejemplo, por Harry Frankfurt en una célebre colaboración a este debate.[135] No estoy seguro de haber comprendido sus argumentos, o en realidad, su versión de esta solución, pero creo que este tipo general de análisis ofrece algunas buenas esperanzas de escape. ¿Tal vez podríamos apelar a razonamientos de punto fijo, para demostrar que por cada persona (o tal vez para

hacerse, y atarse a sí mismo en el universalismo en el desempeño contra el propio juicio clínico de 'ciencia profunda' probablemente sea racional."

[134] Desde luego, hay juegos en que la regresión es verdadera, por lo que no hay solución (no cooperativa) al juego; *véase infra*, sección 2 del presente capítulo, para un análisis de esta cuestión.

[135] "Sin embargo, es posible terminar semejante serie de acciones sin aislarla arbitrariamente. Cuando una persona se identifica a sí misma *decisivamente* con uno de sus deseos de primer orden, este compromiso 'resuena' a lo largo de una disposición potencialmente infinita de órdenes superiores" (Frankfurt, 1971, p. 16). Yo creo que este "efecto de resonancia" debe tomarse como *definición* de lo que significa comprometerse sucesivamente y no como efecto de tal compromiso. En otras palabras, no puedo estar de acuerdo con Frankfurt cuando procede a decir que es "de importancia muy relativa si explicamos esto diciendo que este compromiso implícitamente genera una serie infinita de deseos confirmatorios de órdenes superiores, o diciendo que el compromiso equivale a una disolución del punto de todas las cuestiones concernientes a órdenes superiores del deseo".

algunas personas) existe un nivel más allá del cual la estructura de la preferencia se repite idénticamente? Tengo plena conciencia de que esta sugerencia es a la vez extremadamente oscura y, en la medida en que es inteligible, está preñada de nuevos problemas, pero me propongo dejar la cuestión en este punto.

Como última palabra deseo repetir un punto establecido en el análisis de la norma (iv) en la sección 2 del presente capítulo. Una plena característica de lo que significa ser humano debiera incluir al menos tres rasgos. El hombre puede ser *racional,* en el sentido de sacrificar deliberadamente la gratificación actual por la gratificación futura. El hombre a menudo no es racional, y antes bien muestra *flaqueza de voluntad.* Aun cuando no es racional, el hombre sabe que es irracional y puede *atarse a sí mismo* para protegerse contra la irracionalidad. Esta "segunda mejor o imperfecta" racionalidad se ocupa a la vez de la razón y de la pasión. Lo que se pierde, tal vez, es el sentido de aventura.

III. LA RACIONALIDAD PROBLEMÁTICA: ALGUNOS PROBLEMAS NO RESUELTOS EN LA TEORÍA DE LA CONDUCTA RACIONAL

1. Introducción

El enfoque de la "elección racional" de la conducta humana es, sin mayor duda, el mejor modelo disponible, pero aún quedan por decidir muchos problemas concernientes a su alcance y capacidad de resolver. En este capítulo, presento lo que equivale, en esencia, a una lista de tales problemas. No es una revisión de casos no resueltos en el sentido del discurso de Hilbert de 1900,[1] sino una discusión de anomalías o de "puntos críticos" en el sentido del artículo de 1972 de Oskar Morgenstern.[2] Es decir, no estoy implican-

[1] En el Segundo Congreso Internacional de Matemáticas, David Hilbert presentó una lista de 23 problemas matemáticos no resueltos, algunos de los cuales desde entonces han sido resueltos, otros se ha demostrado que son irresolubles, y otros más quedan por resolver. *Véase también,* Raid (1970).

[2] Morgenstern (1972). Estos "puntos críticos" son los siguientes. *1)* Control de variables económicas. Con esto, Morgenstern se refiere al problema de la racionalidad paramétrica contra la estratégica, analizada en la sección 2 del presente capítulo. *2)* La teoría de la preferencia revelada. *3)* El óptimo de Pareto. Morgenstern no cree que esta noción nos permita evitar las comparaciones interpersonales de utilidades; *véase también,* Morgenstern (1964). *4)* Tâtonnement. Morgenstern se une a los muchos autores que han criticado esta ficción de la economía del equilibrio general. *5)* La Fijación de Walras-Pareto. Aquí, Morgenstern se refiere a la obsesión del economista por

do que éstos sean problemas definidos a los cuales no hay respuestas en la actualidad. Antes bien, que los propios problemas están mal definidos y las soluciones propuestas dependen en mucho de suposiciones *ad hoc*. Lo que se pierde en claridad conceptual por estas soluciones a menudo supera lo que se gana en limpieza técnica.[3]

Tal vez no exista ninguna otra teoría que pueda señalarse como *la* principal competidora del modelo de la elección racional. Una teoría muy general y abstracta de la acción nos permitirá distinguir entre los principales contendientes. Cualquier ejemplo dado de conducta humana puede

la libre competencia, que en opinión de Morgenstern sólo es un caso limitador, sin interés particular. *6)* La asignación de recursos. Aquí Morgenstern critica la profesión económica por no hacer caso de la repercusión de la política sobre la formación y asignación de precios. *7)* Sustitución. Aquí, Morgenstern muestra que en algunos casos la noción de una tasa de sustitución entre bienes en producción no está definida. *8)* Demanda y oferta. Esta sección contiene una crítica de la noción de la demanda agregada. *9)* El análisis de la curva de indiferencia. Aquí Morgenstern muestra que la noción de curvas cerradas de indiferencia alberga una contradicción. También se tratan otras dificultades *10)* La teoría de la firma. Esto es en parte (así como muchos de los otros puntos) un replanteamiento de *1) supra*. *11)* De regreso a Cantillon. Este es un alegato en favor de la disgregación. *12)* La distribución personal y funcional. Esta es una crítica de la teoría de la productividad marginal, y un alegato en favor del uso del análisis de la teoría de los juegos. *13)* Pertinencia de la teoría. Esta sección muestra la correlación inversa entre la importancia empírica de los fenómenos económicos y la intensidad con que se les estudia.

[3] "El único sentido peyorativo legítimo de 'la calidad de ser *ad hoc*' se reduce a una situación en que la eficacia de resolver problemas de una teoría general decrece por virtud de sus dificultades conceptuales cada vez mayores" (Laudan, 1977, p. 117; se suprimieron las cursivas).

ser presentado como el producto final de dos recursos sucesivos de filtración. El primero se define por el conjunto de limitaciones estructurales que reduce el juego de los cursos de acción posibles en abstracto y lo reduce al subconjunto mucho más pequeño de acciones factibles. Se supone que las limitaciones son dadas y no están dentro del control de los agentes; remitimos al lector al capítulo anterior para el caso de las limitaciones autoimpuestas. El segundo proceso filtrador es el mecanismo que elige cuál miembro del conjunto factible se realizará. Las teorías de la elección racional afirman que este mecanismo es la elección deliberada e intencional con el propósito de maximizar alguna función objetiva, ya sea real (como una ganancia) o puramente teórica (como la función de la utilidad que representa las preferencias).

Este modelo de dos pasos inmediatamente sugiere alternativas al enfoque de la elección racional. La primera es la línea de argumentos a la que yo llamaré estructuralista, que niega la *importancia* de la elección racional y en realidad del segundo filtro en general y en cambio afirma la enorme importancia de las limitaciones estructurales. En una versión extrema esto significaría que las limitaciones en conjunto tienen el efecto de reducir el juego factible a un solo punto; en una versión más débil y verosímil, que las limitaciones definen un conjunto tan pequeño que la libertad formal de elección dentro del conjunto no equivale ya a gran cosa. Tanto el estructuralismo en la ciencia social angloamericana,[4] como el

[4] Como ejemplo representativo puede citarse el siguiente texto: "...los estructuralistas subrayan la limitación de la actividad política por la estructura social (incluyendo la clase) y los valores (incluyendo la ideología).

estructuralismo como escuela metodológica francesa [5] pueden embonar, con cierto esfuerzo, en esta definición. Como teoría general de acción me parece demasiado obviamente errónea para merecer gran análisis. En tiempos y lugares en particular, las constelaciones de limitaciones pue-

Los conductistas señalarían la toma de decisiones políticas dentro de estos límites. Los estructuralistas estudiarían la cerca en torno del ganado; los conductistas estudiarían la actividad del ganado dentro de los límites de la cerca". (MacFarland, 1969, p. 135.) Hay aquí algunas ambigüedades. Creo yo que debemos vacilar antes de aceptar la idea de que los *valores* son limitaciones; pertenecen al segundo filtro y no al primero. (Una excepción son los valores lexicográficamente dominantes, analizados en la sección 3 del presente capítulo.) Tampoco estoy de acuerdo con esta opinión ya que concibo al estructuralismo como teórico y no como metodológico. El estructuralista, según mi definición, no es alguien que *decide* estudiar la cerca en torno del ganado antes que los movimientos dentro de la cerca, sino alguien que *afirma* que el ganado tiene muy poca libertad de desplazamiento dentro de la cerca.

[5] Como texto representativo, puede citarse el siguiente: "La résistance que suscite la sociologie lorsqu'elle prétend déposséder l'expérience immédiate de son privilege gnoséologique s'inspire de la même philosophe humaniste de l'action humaine que certaine sociologie qui, en s'armant par example de concepts comme celui de 'motivation' ou en s'attachant par prédilection aux questions de *decision-making*, réalise, à sa maniere, le voeu naïf de tout sujet social: entendant rester maître et possesseur de lui-même et de sa propre vérité, ne voulant connaitre d'autre déterminisme que celui de ses propres déterminations (même s'il leur concede l'inconscience), l'humaniste naïf qu'il y a en tout homme ressent comme une réduction 'sociologiste' ou 'matérialiste' toute tentative pour établir que le sens des actions les plus personnelles et les plus 'transparentes' n'appartient pas au sujet qui les accomplit mais au système complet des relations dans lesquelles et para lesquelles elles s'accomplissent" (Bourdieu *et al.*, 1968, p. 39). [La resistencia que suscita la sociología cuando pretende poseer la experiencia inmediata de su privilegio gnoseológico se inspira en la misma filo-

den ser tales que excluyan todo excepto un pequeñísimo subconjunto de las acciones posibles, pero para que esto pueda ser generalmente cierto tendría que ver algún mecanismo que lo produjera, y en la bibliografía estructuralista nunca he encontrado algún intento de aportarlo. No cabe duda, es cierto, que las clases gobernantes en la mayor parte de las sociedades han tratado de consolidar su dominio reduciendo el conjunto de oportunidades abiertas a las clases oprimidas, pero esta afirmación misma implica que las propias clases gobernantes estaban actuando libre y racionalmente en su propio interés.

Antes de pasar a la segunda alternativa, que es, con mucho, la más importante, deseo señalar una versión más interesante de la primera alternativa. Aun si la conducta en algún caso dado debe explicarse por la teoría de los dos filtros, podremos argüir que sólo el primer filtro es pertinen-

sofía humanista de la acción humana que cierta sociología, la cual armándose por ejemplo, de conceptos como el de "motivación" o adhiriéndose por predilección a las cuestiones de *decision making*, realiza, a su manera, el ingenio devoto de todo sujeto social: pretendiendo seguir amo de sí mismo y de su propia verdad, no deseando saber de otro determinismo que el de sus propias determinaciones (aun si les concede la inconsciencia), el humanista ingenuo que hay en todo hombre reciente como una reducción "sociologista" o "materialista" toda tentativa de establecer que el sentido de las acciones más personales y más "transparentes" no pertenece al sujeto que la realiza sino al sistema completo de las relaciones en las cuales y por las cuales se realizan.] La oscuridad retórica de este pasaje (también representativo en este respecto) hace difícil saber y afirma que las limitaciones son tan fuertes que quitan importancia a la elección; que el segundo filtro opera por medio de un mecanismo distinto del de la elección racional; o que la elección racional de acuerdo con preferencias congruentes no tiene interés comparado al proceso social que genera e internaliza estas preferencias.

te para el análisis de las *diferencias* de conducta. Es decir, podría argüirse que las preferencias y motivaciones de la gente son muy parecidas en todo tiempo y lugar, por lo que las diferencias de la conducta observada deben ser explicadas por diferencias en el conjunto de oportunidades. Las preferencias, por decirlo así, se cancelan unas a otras, dejando las limitaciones estructurales con todo su poder explicativo. Para una ilustración, tómese la variedad de enfoques a la inversión en las sociedades precapitalistas. ¿Se debió la falta de inversión a una carencia de *objetos* de inversión, como lo han argüido, por ejemplo, Douglass North y Robert Thomas,[6] o a la ausencia de la *motivación* de invertir, como lo arguye Eugene Genovese?[7] En general yo creo que los economistas, por alguna variedad de la navaja de Occam, tienden a suponer que las preferencias son casi siempre similares a través del tiempo y del espacio,[8] mientras que los historiadores y los sociólogos son gobernados por su temor a cometer anacronismos e imputar a sociedades premodernas las pautas de preferencia características de ellos mismos. Yo no creo que este hincapié en las limitaciones deba ser llamado estructuralista, y en realidad no representa una alternativa a la teoría de la elección racional. Sin embargo, justifica una aserción en el sentido de que la varianza de conducta puede explicarse por la varianza de oportunidades, lo que en cierto modo es una degradación del enfoque del actor racional.

[6] North y Thomas (1973), p. 62 y *passim*.

[7] Genovese (1965), pp. 16-17.

[8] *Véase* Harsanyi (1955) y Stigler y Becker (1977). Hay aquí una paradoja, en que los más decididos defensores del enfoque del actor racional propongan una teoría que tiene el efecto indiscutible de rebajar la importancia de la elección con comparación con las oportunidades.

La segunda alternativa, a la que he dedicado la mayor parte de este capítulo, cuestiona la *realidad* de la elección racional y afirma que el mecanismo que escoge a un miembro del juego factible es causal y no intencional. Esta alternativa se presenta en dos versiones. La más radical afirma que en cierto número de casos la noción de conducta racional ni siquiera está bien definida; sencillamente no hay una acción exclusivamente prescrita que un actor racional pudiese elegir. En las secciones 2 y 3 del presente capítulo, se defiende este argumento para la importante categoría de los juegos sin solución nocooperativa. En la sección 5 para las situaciones en que la racionalidad conduce a la regresión infinita ya mencionada en la sec. 4, cap. II y en la sec. 9 del presente capítulo para las actitudes mentales contradictorias más extensamente analizadas en el cap. IV. La versión menos radical acepta (o no se compromete acerca de ella) la existencia de una noción bien definida de conducta racional, pero afirma que en algunos casos este ideal no se realiza. En la sección 6 del presente capítulo se seleccionan conductas al azar y conductas tradicionales como los dos principales subcasos de este enfoque. En la sección 5 del presente capítulo analizo un aspecto particular del modelo de la conducta tradicional, a saber, la cuestión de cómo empieza la tradición por establecerse. El surgimiento de la tradición por medio de selección natural, o de una analogía social con tal proceso es el principal mecanismo que se ha propuesto. (Los lectores del capítulo I notarán que aun si este modelo fuese válido en algún caso particular, no habría ninguna presunción de simular la elección racional.)

A estas alternativas podemos añadir una terce-

ra, a saber, el enfoque que niega la *estabilidad* de la elección racional, y que subraya la formación y la transformación de las preferencias. Esta cuestión fue discutida bastante extensamente en la sec. 6, cap. II y en la sec. 8 del presente capítulo. Sólo retorno al problema para establecer unos puntos adicionales que tienen pertinencia para nuestra actual discusión. Los comentarios a la probabilidad subjetiva que aparecen en la sección 4 del presente capítulo también son útiles aquí, ya que yo creo que las teorías de la elección racional basadas en esta noción (es decir, las teorías de la decisión bayesiana) sobrestiman la estabilidad de las probabilidades que pueden ser provocadas por los métodos comunes.

También analizo algunas de las tradicionales objeciones a la racionalidad que no se oyen muy a menudo hoy, pero que, no obstante, conservan cierta fuerza. En la sección 3 del presente capítulo se analiza brevemente y desde el punto de vista de la antropología social y de la teoría de los juegos la noción de las preferencias lexicográficas, que es el contraejemplo más frecuente a la posibilidad de representar preferencias por una función de utilidad. En la sección 7 del presente capítulo es igualmente bosquejada, desde los puntos de vista de la biología, la economía y la sociología la noción de altruismo, que es el contraejemplo habitual a la interpretación que el interés propio da a la racionalidad. Ahora bien, todos sabemos que la elección racional depende de la representabilidad de preferencias o de motivaciones egoístas; sin embargo, yo creo que las soluciones formales a esos problemas no eliminan por completo las cuestiones sustanciales subyacentes.

Concluyo arguyendo que aun si la racionalidad

a menudo no tiene importancia, está ausente o es inestable, sin embargo hay un meollo de casos importantes en que es indispensable el modelo de la elección racional. Afirmo, además, que este modelo es lógicamente anterior a las alternativas, en el sentido de que el científico social siempre debe ser guiado por un postulado de racionalidad, aun si puede terminar descubriéndolo violado en muchos casos particulares. Esta suposición es un "principio de caridad" similar al tan a menudo empleado en la interpretación textual. Nunca deben tomarse en su valor aparente las contradicciones textuales, sino que se debe considerar si el contexto no nos dará una clave para la congruencia. De manera similar, siempre hay que ver muy minuciosamente la conducta aparentemente irracional para ver si, después de todo, surge allí alguna pauta. Huelga decirlo: no hay que extender demasiado la caridad, y puede llegar un punto en que el observador simplemente tenga que afirmar que, hasta donde él lo comprende, la conducta no tiene un motivo racional.

2. Juegos sin soluciones

En esta sección establezco una distinción entre la racionalidad paramétrica y la estrategia.[9] La racionalidad paramétrica suele conducir a una prescripción exclusiva y bien definida para la conducta, dadas las iniciales suposiciones y expectativas. Aquí, el problema es que estas mismas suposiciones y expectativas pueden ser irracionales, por estar fundadas sobre una creencia inge-

[9] En la historia de la ciencia estos enfoques pueden ser relacionados con los nombres de Max Weber y John von Neumann, respectivamente.

nua acerca de un medio constante (o paramétrico). En la racionalidad estratégica esta creencia ya no opera y en cambio entra la visión realista de la naturaleza estratégica del medio. Sin embargo, aquí surge otro problema: en la interacción estratégica puede haber un modo de conducta no exclusivamente definido, al cual se elige como *la* elección racional.

Hay algo extraño o hasta contradictorio en la idea de racionalidad paramétrica. Un agente paramétricamente racional se cree libre de adaptarse óptimamente (dado su fin) a un medio constante, y al mismo tiempo difícilmente no se dará cuenta del hecho de que su medio está integrado en parte por otros agentes similares a él. Desde luego, podría suponer que los otros están tratando de adaptarse óptimamente a lo que a *ellos* les parece un medio constante, y luego adaptarse óptimamente a sus adaptaciones óptimas. Esta es, por ejemplo, la estrategia llamada solución de Stackelberg en el problema del duopolio.[10] Sin embargo, es claro que esto sólo hace que la misma paradoja aparezca a un nivel superior. La contradicción sólo puede ser eliminada por transición al modo de pensar estratégico o de teoría de los juegos. El actor estratégicamente racional se considera participante en un juego, que en el caso ideal es definido como una información perfecta en el sentido de que todos los jugadores tienen un conocimiento cabal de las preferencias y el conocimiento de los demás. Ninguno puede imaginar que está un paso delante de los demás, cada uno debe tomar en cuenta el hecho de que todos toman en cuenta a todos los demás antes de tomar su decisión. Esto parece como una re-

[10] *Véase* Coddington (1968), pp. 58 *ss.*; también Elster (1978a), cap. v.

gresión infinita, pero como se observó al final del capítulo II, es posible, mediante la noción de un *punto de equilibrio*, causar un corto circuito en la regresión y llegar a una solución exclusivamente definida para el juego. Un punto de equilibrio es un conjunto de estrategias individuales, cada una de las cuales es óptima contra las demás. En otras palabras $(s_1 \ldots s_n)$ es un punto de equilibrio si para todo i, s_i es la mejor respuesta del jugador i a $(s_1 \ldots s_{i-1}, s_{i+1} \ldots s_n)$. (Pronto volveré a una ambigüedad fundamental en la palabra "mejor" aquí.)

La idea de racionalidad paramétrica, cualesquiera que sean sus fallas, tiene al menos la ventaja de que suele estar bien definida. Ciertamente, hay problemas de maximización sin solución, como cuando la función objetiva es discontinua o el conjunto factible no es compacto. Un caso importante es la no existencia de planos óptimos en ciertos problemas de planeación para un horizonte infinito.[11] Sin embargo, en la mayor parte de las aplicaciones, la suposición de un medio constante genera una acción exclusiva de *lo* racional que debe hacerse, o al menos un conjunto de acciones las cuales son igual y máximamente buenas. Estas acciones pueden resultar contraproducentes si se basan en suposiciones sobre otros agentes que son incompatibles con las suposiciones de estos agentes acerca de sí mismos, pero en todo caso están bien definidas. A este respecto, la racionalidad estratégica es mucho más problemática, dada la importancia omnipresente de los juegos sin una solución no cooperativa. En esta sección analizo dos categorías de tales juegos: los juegos de suma variable sin

[11] Heal (1973), cap. XIII, tiene una buena discusión; *cf.* también cap. II, n. 62, *supra*.

puntos de equilibrio y de suma variable con varios puntos de equilibrio, ninguno de los cuales puede ser elegido como la solución. En la sección 3 del presente capítulo se menciona una tercera categoría, los juegos de suma constante sin una solución de Von Neumann-Morgenstern.

Un juego sencillo sin un punto de equilibrio es el siguiente: "Cada jugador escribe un número, el jugador que ha escrito el número más grande recibe de cada uno de los demás jugadores una suma de dinero igual a la diferencia entre el número mayor y el número anotado por este jugador." Es intuitivamente claro, y fácil de probar,[12] que en esta *fuga hacia adelante* no puede haber punto de equilibrio. La economía de la hiperinflación tal vez pueda interpretarse como un juego de esta forma general. En este caso no hay punto de equilibrio porque el conjunto de estrategia no es compacto, y el fenómeno probablemente no es más importante que los casos en que la racionalidad paramétrica tampoco logra definir una conducta exclusivamente caracterizada.

Mucho más importantes son los juegos sin solución en que hay *varios* puntos de equilibrio. John Harsanyi ha ofrecido recientemente un análisis poderoso de este problema,[13] afirmando que en realidad se puede definir una solución para todos los juegos no cooperativos. No es posible entrar aquí en los detalles de su argumento, pero sí caben algunas observaciones críticas. Primera y más obvia, su concepto de solución no satisface los requerimientos ordinarios de conducta racional al menos para algunos tipos de juegos. De hecho, para un juego como el siguiente, Harsanyi afirma que los jugadores deben utilizar sus

[12] Owen (1968), p. 73.
[13] Harsanyi (1977), caps. VII y XIV.

estrategias maximin, aun si no son óptimas unas contra otras.

	B_1	B_2
A_1	30, 20	20, 0
A_2	40, 0	10, 30

Comentando este argumento, Harsanyi reconoce que "puede argüirse que esta sugerencia no merece ser llamada 'solución', porque carece de muchas de las deseables propiedades de las primeras soluciones. Se le puede llamar una 'casi solución'. Hasta se podría argüir que es simplemente otro modo de decir que tales juegos *no* tienen verdaderas soluciones".[14] Esta es, en realidad, mi opinión. La estrategia maximin en tales casos no debe llamarse estrategia racional, pues ello significaría que si los otros jugadores son racionales, entonces es irracional el ser racional. El uso de la estrategia maximin (o de *cualquier* estrategia) equivale a levantar los brazos al cielo y buscar una manera de salir de la regresión infinita simplemente porque hay que *actuar*. Sin embargo, la acción es un resultado de presiones causales que se forman y de fatiga mental, no de deliberación racional.

En segundo lugar (y mucho más controvertible) no es seguro que el concepto de solución de Harsanyi sea satisfactorio en los casos en que sí elige un punto de equilibrio como la solución. Uno de los rasgos importantes del libro de Harsanyi es el hincapié que hace en los problemas conceptuales inherentes a las soluciones que incluyen estrategias mixtas. Es bien sabido que en un punto de equilibrio de estrategias mixtas, la estrategia mixta s_i nunca podrá ser inequívoca-

[14] *Ibid.*, p. 138.

mente *la* mejor para el jugador i contra $(s_1 \ldots s_{i-1}, s_{i+1} \ldots s_n)$, pues igualmente bien le irá si utiliza cualquiera de las estrategias puras que entran en s_i, o en realidad cualquier mezcla de probabilidades de estas estrategias. Esto demuestra una carencia de estabilidad de tales puntos de equilibrio que resulta un tanto perturbadora, especialmente de un jugador cauteloso que preferiría utilizar su estrategia maximin si esto le asegura el mismo rendimiento contra las estrategias de equilibrio de los otros jugadores. Harsanyi resuelve este problema eligiendo una estrategia como la privilegiada, es decir, la estrategia mixta *centroide* en que todas las estrategias puras tienen la misma posibilidad de ser escogidas. Esto implica que las estrategias mixtas pueden entrar en la solución de un juego sólo si son equiprobabilísticas en este sentido. Este procedimiento es interpretado causalmente, y no intencionalmente; es decir, Harsanyi presupone que estas probabilidades no serán deliberadamente generadas por un recurso mecánico, como echar una moneda al aire, "sino antes bien por lo que equivale a un mecanismo aleatorio inconsciente, dentro del sistema nervioso central por i".[15] En la sección 6 del presente capítulo, volveré brevemente a algunas de las implicaciones conceptuales de este enfoque. Aquí me limitaré a afirmar que no es obvio que se conforme a nuestras nociones intuitivas acerca de la racionalidad. Para elucidar este punto, considérese el siguiente juego:

	B_1	B_2
A_1	2, 2	1, 2
A_2	2, 1	1, 1

[15] *Ibid.*, p. 114.

Como observa Harsanyi,[16] esto es, en un sentido, una variante del Dilema del preso. Para cada jugador es absolutamente indiferente si escoge su primera o su segunda estrategia; recibe el mismo nombramiento en ambos casos, sin que importe lo que haga el otro. También podría elegir cualquier estrategia mixta con el mismo resultado esperado. No obstante, es claro que va en ventaja de los dos jugadores llegar a (A_1, B_1), y no a (A_2, B_2), según Harsanyi, la estrategia mixta centroide es la solución a este juego, con un aumento esperado de 3/2 para cada jugador. Yo supongo que este resultado es contraintuitivo. Sin duda, la idea de una solución debe ser construida en tal forma que capte nuestras nociones intuitivas acerca de la racionalidad; si la intuición nos dice que los jugadores racionales elegirían (A_1, B_1) y la teoría de Harsanyi nos dice otra cosa, tanto peor para la teoría. No quiero ser dogmático, aquí, acerca de la intuición. En muchos casos nuestras intuiciones son bastante endebles,[17] y

[16] *Ibid.*, p. 278 *ss.*

[17] Considérese, por ejemplo, el siguiente juego *(ibid.*, página 287):

		Jugador II		
		B_1	B_2	B_3
Jugador	A_1	6, 4	0, 0	0, 0
	A_2	0, 0	4, 6	0, 0
	A_3	0, 0	0, 0	2, 3

Aquí, la teoría de Harsanyi prescribe que (A_3, B_3) sea la solución. Esto parece estar en armonía con algunas de nuestras intuiciones acerca del juego, aun si tal vez entra en conflicto con otras intuiciones. El papel del análisis conceptual consiste en fijar tales casos limítrofes, aunque conformándose (de la manera más sencilla posible) a nuestras intuiciones inconmovibles. Puede ocurrir, como lo ha dicho John Rawls (1971), que antes de que

entonces han de conformarse a la teoría y no al revés. Sin embargo, hay algunas intuiciones que son difíciles de abandonar, y yo creo que esta es una de ellas. El hecho de que estudios experimentales [18] muestren que las segundas estrategias que son elegidas al menos por un jugador en cerca de 20 % de los casos probablemente no es prueba contra la confiabilidad de esta intuición. Como ya lo observó Rapoport, la defección también puede explicarse suponiendo que para algunos jugadores la diferencia entre sus propios rendimientos y los de los oponentes también entran en el cálculo, lo que significa, desde luego, que la matriz de rendimiento no refleja en realidad los rendimientos de utilidad de los jugadores. En una situación experimental, los rendimientos deben darse en dinero o en términos similarmente materiales, y aun así existe siempre la posibilidad de que los efectos de interacción hagan que la verdadera estructura de rendimiento diverja de la material. *Si* la verdadera estructura de rendimiento es como se indica en el juego, yo creo que la elección de (A_1, B_1) es inequívocamente el resultado racional que luego podrá surgir como limitación para todos los posibles conceptos de solución.

Las consideraciones anteriores implican que al menos en algunos casos una situación similar a un juego no ofrece una conducta racional exclu-

lleguemos a un "equilibrio reflexivo" algunas de nuestras intuiciones, hasta las que parecían absolutamente firmes, tengan que ser modificadas. No excluyo la posibilidad de que, pensándolo mejor, las teorías de Harsanyi lleguen a parecer tan atractivas que yo tendré que modificar mi intuición acerca de lo que un jugador racional haría en la variante del Dilema del preso presentada en el texto, pero en la actualidad no veo ninguna razón para hacerlo.

[18] Rappoport *et al.* (1976), pp. 226-227.

sivamente definida. Casos bien conocidos tomados de la bibliografía de la teoría de los juegos son la "gallina" [19] y la "batalla de los sexos",[20] en que un estancamiento en las negociaciones impide que surja una solución no cooperativa. Un ejemplo, basado en una situación generalizada de "gallinas" puede ser útil en este punto. En teoría económica existe un conocido problema concerniente al eslabón entre la escasez de mano de obra y la tendencia del factor en el progreso técnico. Desde Marx hasta Hicks [21] se dio más o menos por sentado que los empresarios tienen un incentivo en economizar los factores escasos, no sólo mediante sustitución sino también mediante una tendencia en sus innovaciones. En una contribución justamente célebre, W. E. G. Salter [22] mostró entonces que después de ocurrida la sustitución, todos los factores son igualmente escasos en el margen, de modo que los empresarios no tienen un incentivo especial en sesgar su búsqueda de un modo, y no de otro. Se emplearon dos líneas de argumento para responder a esta afirmación. Una de ellas está vinculada a los nombres de Kennedy, Von Weiszäcker y Samuelson,[23] y dice en esencia que hay limitaciones al punto de innovaciones posibles que hacen racional la tendencia. La otra fue propuesta por William Fellner,[24] quien arguyó que era racional que los empresarios sesgaran sus innovaciones si esperaran que el alza de salarios continuara en el futuro. Este último argumento ha sido criticado por

[19] Rapoport (1966), pp. 137 *ss.*; también extensamente analizado en Rapoport *et al.* (1976).
[20] Luce y Raiffa (1957), pp. 90 *ss.*, 115 *ss.*
[21] Esp. Hicks (1932), p. 129.
[22] Salter (1960), pp. 43-44.
[23] Notablemente, Kennedy (1964).
[24] Fellner (1961).

Paul David,[25] por motivos de que descuida la importancia de los personajes solitarios en el cambio técnico. Yo mostraré que el argumento de David, una vez elaborado, muestra que el juego al que juegan los empresarios es un juego sin solución.

El razonamiento en que se apoya la teoría de Marx-Hicks debió de ser, aproximadamente, el siguiente. Es benéfico para todos los capitalistas si todos introducen inovaciones que ahorren mano de obra, pues entonces se reducirá la demanda agregada de mano de obra, y con ella, los salarios. Por tanto, puede esperarse que un aumento de salarios produzca innovaciones que ahorren mano de obra en tal forma que reduzcan los salarios a su nivel anterior. Este razonamiento habría sido válido si los capitalistas actuaran como un actor, pero falla cuando se aplican las decisiones descentralizadas. Si todos los demás capitalistas van a sesgar sus innovaciones, el capitalista individual puede actuar como hombre libre; si no lo hace, su acción puede tener sólo un efecto infinitesimal sobre el nivel de salarios, suponiendo una competencia perfecta. Ahora bien, si se espera que el nivel de salarios aumente en el futuro, es racional que el empresario individual sesgue su investigación, como lo arguyó Fellner, porque esto le permitirá prevenir hasta cierto grado los futuros aumentos de salario. Sin embargo, la pregunta es por qué el empresario debe tener estas expectativas. La única base racional de tales expectativas debe ser que los otros empresarios no sesgarán su investigación sino, antes bien, continuarán actuando como antes: es decir, una suposición acerca de un medio paramétrico. Sin embargo, resulta difícil llamar racional a esta

[25] David (1975), pp. 36 n. 2, 54.

suposición. Sea como fuere, el empresario pronto olvidará esta suposición cuando resulte que los salarios bajan en lugar de subir, porque todos los empresarios —y no sólo él— han introducido inventos que ahorran mano de obra. El siguiente paso puede ser, ya que nadie introduzca esta tendencia porque todos esperan que los otros lo hagan, ya que finalmente todos comprendan que la situación es estratégica. En el último caso, tenemos uno generalizado de "gallinas": si todos los demás introducen innovaciones que ahorran mano de obra, entonces es racional que el individuo no lo haga, pero si nadie más lo hace, entonces va en su interés hacerlo.

En esta situación, como se explicó antes, no hay conducta que sea la *racional*, al menos no si estamos hablando acerca de racionalidad estratégica y no paramétrica. Mientras los empresarios crean que cada uno, individualmente, está actuando en un medio de otros empresarios que deben seguir actuando como lo hicieron el año anterior, entonces tiene sentido la noción de la conducta individualmente racional, para esta creencia. Sin embargo, la creencia misma no resulta muy sensata. Y en cuanto la creencia es suplantada por la visión más racional de la naturaleza estratégica y simétrica de la situación, la idea de lección racional simplemente se evapora, y la conjetura tiene que suplantarla.

Supongo, pues, que algunos de tales juegos sin solución constituyen una profunda anomalía en la conducta de la teoría racional. Si todos los juegos tuviesen una solución no cooperativa, entonces el actor estratégicamente racional sería la perfecta encarnación de la mónada leibniziana. No tendría "ventanas", en el sentido de no poder comunicarse con los demás, y sin embargo su

punto de vista sobre el universo reflejaría e internalizaría todos los demás puntos de vista en un todo armónico. El hecho de que esta tácita coordinación o preestablecida armonía no siempre sea posible, y que *no puedan* hacerse algunas elecciones por el ego racional, tiene implicaciones para la filosofía, la psicología y las ciencias sociales en general.

3. Preferencias lexicográficas

Esta estructura de preferencia es el contraejemplo habitual que muestra que no todos los ordenamientos transitivos y completos de preferencia pueden ser representados por una función de la utilidad de valor real. Un vector $(a_1 \ldots a_n)$ es lexicográficamente preferido sobre otro $(b_1 \ldots b_n)$ si y sólo sí para algunos i, $a_i > b_i$ y para todos $j < i$, $a_j = b_j$. Una implicación de esta definición es que para todo $j > i$, la relación entre a_i y b_i es totalmente improcedente en la evaluación de los vectores. En particular, b_j puede ser mucho mayor que a_i para todo $j > i$, y sin embargo el vector a será lexicográficamente preferido a b si a_i es ligeramente mayor que b_i y $a_i = b_i$ para $i < j$. O, en otras palabras, para cada i, el i ésimo componente del vector es incomparablemente más importante que el $i + 1$ ésimo de modo que no puede haber cambios entre los dos. Es bien sabido que esta estructura de preferencia no puede ser representada por una función ordinal de utilidad,[26] ni por una de Neumann-Morgenstern.[27] El último

[26] *Véase*, por ejemplo, Luce y Suppes (1965), pp. 261 *ss.*

[27] Riker y Ordeshook (1973), pp. 37 *ss.* Podemos notar aquí que si el número de alternativas es finito, las preferencias siempre pueden ser representadas por una función de utilidad ordinal, mientras que las preferencias lexico-

hecho también tiene implicaciones para la teoría de los juegos. El teorema minimax, según el cual todos los juegos finitos de suma cero, de "lo que se gana se pierde", tiene una solución en las estrategias mixtas presupone, fundamentalmente, que las preferencias están representadas por funciones de utilidad que son únicas hasta una transformación lineal positiva. Por tanto, podemos conjeturar que si los jugadores tienen preferencias no representables, el juego de "lo que se gana se pierde", puede no tener solución. Peter Fishburn [28] ha demostrado que esto es en realidad lo que ocurre. Para ser precisos, ha demostrado que en los juegos de "lo que se gana se pierde" con preferencias lexicográficas hay *1)* casos en que el jugador I tiene una estrategia maximin del jugador II una estrategia minimax, pero éstas no son óptimas una contra otra y *2)*, casos en que ninguno de los dos jugadores tiene una estrategia maximin o minimax. Este último caso queda ilustrado por el siguiente juego, en que las ganancias para el jugador I están representadas por preferencias lexicográficas tridimensionales:

		Jugador II		
		s_1	s_2	s_3
Jugador I	t_1	1, 0, 0	0, 0, 0	0, 1, 1
	t_2	0, 0, 0	0, 1, 1	0, 1, 0

gráficas no pueden ser representadas cardinalmente ni aun en este caso. Por esta razón, la imposibilidad de la representación cardinal tal vez sea el más interesante de estos teoremas. También podemos observar que los ordenamientos lexicográficos son tal vez más frecuentes en la vida política que en la vida económica; *véase* también Riker y Ordeshook (1973) y esp. M. Taylor (1973).

[28] Fishburn (1973).

Suponiendo que I escoge t_1 y t_2 con probabilidades a y $(1-a)$, y que II escoge s_1, s_2, s_3, con probabilidades p, q, r, entonces el rendimiento tridimensional esperado de I es $(ap, r+(1-a)q, ar+(1-a)q)$. Si $a=0$, entonces lo peor que puede ocurrir a I es $(0, 0, 0)$, a saber si II fija $p=1$, $r=q=0$. Si $a>0$, entonces el peor resultado posible para I es $(0, 1-a, 1-a)$, para $p=r=0$, $q=1$. Pero esto significa que I no puede elegir a de tal modo que maximice este resultado mínimo. Para cada posible a hay una garantía más pequeña que le asegura un mínimo mejor, pero ir al límite y equiparar a con 0 no produce el mínimo mejor.

¿En qué estriba la importancia de estos resultados formales? Es fácil —y hasta cierto punto, justificado— desdeñarlos, tildándoles de improcedentes para la teoría de la conducta racional. Hasta donde yo sé, nadie ha ofrecido todavía un ejemplo interesante de un juego de "ganar o perder todo" de la vida real, con preferencias lexicográficas, y en todo caso no parecen ofrecer ningunas paradojas que no se encuentren también en los agentes paramétricamente racionales con funciones objetivas discontinuas o conjuntos y no conjuntos. En cuanto a la no representabilidad de las preferencias, esto no muestra que la conducta lexicográficamente gobernada no pueda ser reductible al análisis de selección natural, sólo que este análisis no puede utilizar la cómoda herramienta de la función de utilidad. No obstante, sí pueden discernirse algunas cuestiones sustanciales tras las nubes de confusión que parecen surgir en toda discusión de estas cosas.

La primera concierne a la interpretación de los valores lexicográficamente primeros como limitaciones a la toma de decisiones, y no como normas

para la toma de decisiones. Desde luego, podemos decir que al nivel de subsistencia, las calorías son lexicográficamente más importantes que el sabor,[29] pero una interpretación más natural podría ser, más a menudo, decir que la limitación de las calorías debe añadirse a la limitación presupuestaria, para la definición de un conjunto factible. Esto da cierta fuerza al argumento de que la elección racional debe ser *sin importancia* y que las limitaciones estructurales que definen el conjunto factible son más decisivas. La aversión marxista a las teorías de la elección del consumidor pueden explicarse en parte por el hecho de que Marx escribió en una época en que la limitación del presupuesto y la limitación de las calorías tenían, en conjunto, el efecto de estrechar el conjunto factible, hasta ser, en realidad, un muy pequeño subconjunto, aun cuando variaciones regionales del consumo durante la Revolución Industrial mostraron que era posible vivir al nivel de subsistencia de varias maneras distintas.[30] (La aversión puede explicarse en parte por la suposición marxista de que las preferencias del consumidor, en el capitalismo, son sumamente manipulables y, por tanto, *inestables.)*

La segunda cuestión es más importante. Se refiere a la relación entre "conducta racional" y "conducta económica". A este respecto podemos citar dos ideas opuestas:

> He llegado a creer que el enfoque económico es muy comprensivo, y que es aplicable a toda conducta humana, ya sea a una conducta que incluya precios monetarios, o imputados precios de som-

[29] Georgescu-Roegen (1954).

[30] A. J. Taylor (1960); para las implicaciones de este factor para la teoría del valor trabajo, *véase* Morishima y Seton (1961) y Elster (1978c).

bras, de listas repetidas o infrecuentes, decisiones grandes o pequeñas, fines emocionales o mecánicos, personas ricas o pobres, hombres o mujeres, adultos o niños, personas brillantes o estúpidas, pacientes o terapeutas, hombres de negocios o políticos, profesores o estudiantes.[31]

Una suposición básica... es que cierto tipo de "intercambio" siempre será posible. Formalmente podemos expresar esto suponiendo el llamado *Axioma de Arquímedes*. Esto significa en nuestro ejemplo que si tenemos

$$(x_1, y_1) \text{ de preferencia sobre } (x_2, y_2)$$

siempre podemos invertir la preferencia aumentando y_2; es decir, siempre existe un $y > y_2$ de modo que

$$(x_2, y) \text{ de preferencia sobre } (x_1, y_1).$$

Esto significa que una pérdida de algunas unidades de un artículo siempre puede ser compensada por un aumento de algunas unidades de otro artículo o, dicho de otra manera, *todo tiene su precio*. Puede ser tentador definir la economía como la ciencia de las cosas que tienen un precio, en un sentido muy general. Las cuestiones de vida o muerte y los principios éticos como una aversión absoluta al juego de apuestas serían considerados, entonces, como pertenecientes a las ciencias sociales más generales.[32]

Yo supongo que la segunda opinión es la más razonable, precisamente por el hecho de las preferencias no-arquimédicas (de las cuales la preferencia lexicográfica es un caso especial). Cuando no se satisface la norma arquimédica, nos enfrentamos a artículos o actividades que en cierto sen-

[31] Becker (1976), p. 8.
[32] Borch (1968), p. 22.

tido no son comparables ni se prestan al enfoque económico. Desde luego, esto no significa que no se les pueda reducir al análisis racional. La incapacidad del enfoque económico para explicar ciertas formas de conducta no implica que debamos invocar limitaciones estructurales, la tradición, la selección natural o el azar; pueden ser una conducta racional no económica. Creo yo que algunas cuestiones controvertidas en la antropología social pueden replantearse en este marco. El debate de las definiciones de economía "sustantiva" contra "formal" [33] puede interpretarse como una discusión de preferencias no arquimédicas contra arquimédicas. Donde las preferencias tienen la propiedad arquimédica, podemos hablar como si la gente maximizara la utilidad a lo largo de una sola dimensión, de otra manera, hay que reconocer explícitamente la jerarquía de los valores. El análisis de las "esferas económicas" [34] también se relaciona con este problema. Cuando el intercambio se organiza en esferas jerárquicamente dispuestas con diferentes medios de cambio, nos enfrentamos a un sistema no arquimédico. Para un ejemplo similar tomado de nuestras propias sociedades, se supone que las medallas "no tienen precio".

Se puede plantear la pregunta de si estas jerarquías existen en realidad o sólo en la ideología. Fredrik Barth arguye que las esferas económicas no son tan herméticas como parecen ser,[35] y nuestras propias sociedades es bien sabido que al menos algunas distinciones honoríficas están, extraoficialmente, a la venta (una excepción impor-

[33] *Véanse* los artículos reunidos en Le Clair y Schneider (comps.) (1968).
[34] *Véase*, por ejemplo, Bohannan (1955).
[35] Barth (1967).

tante son los honores militares. "Je ne vois que la condamnation à mort qui distingue un homme, pensa Mathilde. C'est la seule chose qui nes'achète pas".)[36] La función empresarial, subrayada por Barth, es facilitar el movimiento entre las esferas que en la ideología oficial están totalmente separadas, probablemente se encuentra por doquier. No obstante, las ideologías también son parte de la realidad, y la disimulada transferencia de artículos a través de las barreras que hay entre las esferas es un proceso muy diferente del fluir automático de artículos en las economías modernas.

4. La probabilidad subjetiva

Los dos pilares de la moderna teoría de la decisión son la utilidad cardinal y la probabilidad subjetiva. Ciertamente, es un triunfo formal de la filosofía y de la economía el haber relacionado estas nociones subjetivas con normas operativas y observables. Dando al sujeto una serie de elecciones entre opciones riesgosas, es posible definir conceptos que expresen las utilidades cardinales asociadas por él a ciertos hechos así como la probabilidad subjetiva con que cree que ocurrirán.[37] La pregunta es, desde luego, si esta operación en realidad mide lo que se propone medir; si hay distorsiones inherentes al procedimiento que impiden ser válidas a estas técnicas. No analizaré aquí el problema de la utilidad cardinal, salvo para mencionar la objeción común de que es inapropiado que la actitud del individuo que se arriesga deba ser pertinente a la medición de la

[36] Stendhal (1952), I, p. 489.

[37] Luce y Suppes (1965) ofrecen una revisión completa del campo.

utilidad que atribuye a *ciertos* resultados.[38] En cambio, afirmaré que la noción de probabilidad subjetiva es menos útil para una teoría de la toma racional de decisiones de lo que suele argüirse de la filosofía bayesiana, y que a menudo es más racional reconocer ignorancia que esforzarse por alcanzar una cuasi-precisión numérica en la medición de creencias. Si se fijan demasiado altas las normas de racionalidad, de ello puede derivar la irracionalidad.

El enfoque bayesiano —utilizar probabilidades subjetivas— a la conducta y la elección humanas, está siendo usada cada vez más por economistas y filósofos por igual.[39] Una de sus implicaciones principales es eliminar la norma maximin en decisiones en caso de incertidumbre (es decir, total ignorancia). En primer lugar, se nos dice, no existe algo que pueda llamarse ignorancia total; y en segundo lugar, el único uso para la norma maximin está en la solución de juegos de ganar todo o perder todo, donde el medio (es decir, el otro jugador) puede predecirse con total certidumbre. En lugar de la norma maximin se recomienda, entonces, emplear la norma de la utilidad esperada, o que maximicemos la utilidad esperada sobre la base de alguna distribución de probabilidades definidas sobre el conjunto de alternativas. Desde luego, nadie ha negado que ésta es una buena norma en muchos casos. La novedad del enfoque bayesiano está en que arguye que se le *debe* usar siempre porque siempre es posible definir un con-

[38] Arrow (1963), p. 10 y Sen (1970), p. 97; pero *véase* también Morgenstern (1976).

[39] Así Harsanyi (1975), arguyendo contra John Rawls; también Harsanyi (1977), Raiffa (1968), Keeney y Raiffa (1976).

junto de probabilidades subjetivas que pueda servir como guía para la acción. En esta sección, yo deseo cuestionar este paso de "es" a "debe ser".

Dados un acontecimiento E y un agente X siempre *es* posible diseñar una serie de cuestiones que obtengan de X un número que pueda interpretarse como su asignación de probabilidad subjetiva a E. Comenzaremos, por ejemplo, dándole la elección entre las opciones siguientes:

O_E: Si E ocurre, usted obtiene una recompensa R. Si E no ocurre, no obtiene usted nada.

O_1: Si se saca, al azar, una bola roja de una urna que contenía el porcentaje p_1 de pelotas rojas y el porcentaje $100 - p_1$ de pelotas negras, usted obtiene la recompensa R. Si saca una bola negra, no obtiene usted nada.

Suponemos que X prefiere O_E a O_1. Si suponemos que maximiza la utilidad esperada, debemos concluir que asigna a E una probabilidad subjetiva mayor que p_1. Podemos entonces repetir el mismo experimento con O_2 en lugar de O_1, donde O_2 difiere de O_1 sólo en que p_1 es remplazada por la probabilidad $p_1 + e$. Si el agente aún prefiere O_E a O_2, lo colocamos entre la elección de O_E y O_3, donde la probabilidad es $p_1 + 2e$, etcétera. Tarde o temprano llegaremos a una lotería O_n que X prefiere a O_E, de modo que su probabilidad subjetiva es menor a $p_1 + (n - 1)e$. Volvamos entonces a elegir entre O_E y O_{n+1}, donde ahora utilizamos la probabilidad $p_1 + (n - 3/2)e$. Continuamos de esta manera todo lo que se desee, y definimos la probabilidad subjetiva con cualquier grado de precisión.

Este es sólo uno y el menos refinado de los muchos métodos que pueden emplearse para pro-

vocar probabilidades subjetivas o grados de fe. No obstante, muestra bastante bien la *posibilidad de histéresis*[40] que es el punto central que aquí plantearemos. ¿Cómo podremos estar seguros de que la probabilidad final $\bar{p}$ es independiente del camino que ha conducido a ellas, es decir, independiente de los valores particulares de p_1 y de e que fueron elegidos? Deseamos que $\bar{p}$ no refleje nada más que la información pertinente (y toda ella) que el individuo tiene en su posesión con respecto a E; en otras palabras, queremos que $\bar{p}$ sea una *variable de estado*[41] que describa cierto rasgo del individuo en cierto momento. Sin embargo, si el método para medir el valor de esta variable tiene cierto efecto sobre el valor que medimos, entonces, obviamente, la interpretación de $\bar{p}$ debe ser muy distinta. En este caso, $\bar{p}$ también depende de procesos causales que son improcedentes desde el punto de vista de la información o de la creencia subjetiva. Existe una buena evidencia experimental de que tales distorsiones debidas al *anclaje* surgen en muchos casos.[42] (Además, hay distorsiones debidas a varias formas de racionalización de los deseos,[43] así como el sutil fenómeno del autoengaño, analizado en la sec. 4, cap. IV.) Desde luego, estos resultados no dependen de una comparación entre las probabilidades subjetivas provocadas y las verdaderas probabilidades, pues este último concepto no es operativo. Antes bien, de-

[40] Elster (1976a).

[41] *Véase* Fararo (1973), pp. 200 *ss.*, 487 *ss.* La definición de un estado dado por teóricos de sistema (por ejemplo: Padulo y Arbib, 1974, p. 21) es demasiado general para nuestros fines actuales; *véase* Elster (1976a) para otra discusión.

[42] Tversky y Kahneman (1974), pp. 1128 *ss.*

[43] *Véase*, en particular, Janis y Mann (1977).

penden de comparaciones entre diversos métodos para provocar las probabilidades subjetivas, métodos que deben ser equivalentes pero, a pesar de todo, producen evaluaciones diferentes.

La idea es que debemos descartar los enunciados de creencia que sean resultado de un proceso causal; todos los enunciados de creencias son causalmente producidos. La idea es que debemos excluir los enunciados de creencia que son generados por procesos causales *improcedentes a la creencia*, es decir, procesos causales en que las causas incluyen no sólo *razones* sino también factores causales improcedentes, como fatiga mental (que es mayor cuanto más bajo sea p_1 y más pequeña sea e), una preferencia por la simetría (¿por qué no 50 %?) o una aversión a la simetría (sin duda, 50% es demasiado simplista). En la sec. 3 del cap. II, hemos analizado con cierto detalle un caso de tales procesos causales improcedentes a las creencias que condujeron aún a la formación de creencias, a saber, la apuesta de Pascal. Uno de los puntos establecidos allí fue que no debemos aferrarnos a una creencia, o actuar de acuerdo con ella si sospechamos que es el resultado de procesos improcedentes. Si yo tengo una tendencia periódica a las alucinaciones *y lo sé*, entonces debo tener cuidado de no actuar como si mis creencias fueran ciertas. (O si las alucinaciones van con una falta de juicio a este respecto, debo precaverme, de antemano, de no actuar de acuerdo con ella.) Presupongo que casi lo mismo es válido para las posibilidades subjetivas provocadas por un procedimiento como el que antes especificamos. Supongo además que ningún método para mejorar este procedimiento podría evitar esta dificultad. Aquí, por lo general la sagacidad vale de poco,

como ya se indicó: si mostramos a un sujeto su preferencia irracional o improcedente por la simetría, podrá reemplazarlas por una aversión, no menos distorsionante, a la simetría.

Entre los teóricos de la decisión que han abrazado opiniones parcialmente similares debemos citar en especial a William Fellner.[44] Fellner arguye que si las probabilidades subjetivas son interpersonalmente inestables puede ser racional sesgarlas en una u otra dirección para evitar o reducir el arrepentimiento y las autorrecriminaciones que pueden seguir a actuar de acuerdo con tales probabilidades. Surge entonces esta pregunta: ¿hasta qué punto sesgarlas, y por qué razones? Si las razones son las antes aducidas, que "el grado de calibración depende del procedimiento de provocación",[45] entonces no hay argumento para parar, que no sea el conservadurismo más completo y el uso de la norma maximin. Sin embargo, no es esto lo que Fellner estaba pensando. Si las razones son que nos encontramos ante una decisión irreversible, Claude Henry ha mostrado que puede ser racional no actuar de acuerdo con la estimación actual de las probabilidades subjetivas, aun suponiendo que no fuesen deformadas por histéresis.[46] Sin embargo, en el argumento de Fellner no encontramos nada tan específico. Su razón sencillamente parece ser el deseo de evitar la incomodidad subjetiva que podemos sentir cuando se nos demuestra que estamos equivocados en el sentido *ex post facto*. Desde luego, es bien sabido que los burócratas y otros que toman decisiones pueden tener tendencias en su decisión, por este mecanismo, por-

[44] Fellner (1965).
[45] Tversky y Kahneman (1974), p. 1130.
[46] Henry (1974).

que la culpa inseparable de una conjetura errada es mucho más grande que el elogio dado por una atinada. Sin embargo, yo dudo mucho de que ésta sea una práctica que debamos fomentar.

De estas observaciones, generalmente escépticas, podemos sacar algunas conclusiones. Una posibilidad es proponer el uso de razonamiento maximin en tales casos. Si yo no confío en mis probabilidades subjetivas y no tengo un fundamento teórico para un procedimiento que pudiese corregir mis probabilidades provocadas, de modo que estuviesen más cerca de mis "probabilidades subjetivas reales", ni para un procedimiento que pudiese definir un intervalo dentro del cual se encontraran esas probabilidades reales, entonces sin duda estaré en un estado de total y auténtica ignorancia y deberé adoptar la norma maximin o una de las otras normas congruentes con la caracterización que Arrow-Hurwicz dan de la ignorancia total.[47] Las cuestiones hoy tan debatidas de la energía nuclear y del ADN recombinante parecen caer en esta categoría. Aun suponiendo que los riesgos de falla del reactor pudieran ser numéricamente calculadas,[48] aun quedan factores políticos[49] y geológicos,[50] en que parece prevale-

[47] Arrow y Hurwicz (1972); también Luce y Raiffa (1957), pp. 296 *ss.* La caracterización implica que sólo las mejores posibles y las peores posibles consecuencias son pertinentes para la elección entre las alternativas.

[48] Esto tal vez sea conceder demasiado. En particular, el problema del error humano incluye no sólo fallas de modo común, sino también efectos ulteriores que pueden ser difíciles de calcular. (*Véase* Feller, 1968, pp. 118 *ss.*, para la distinción entre estos dos problemas.)

[49] En el Instituto de Investigación de Stanford (1967) encontramos, por ejemplo, probabilidades cuantificadas para los intentos de sabotear plantas de energía nuclear, que difícilmente podemos tomar en serio. Sea como fuere, el plutonio legado a generaciones posteriores seguirá

cer una auténtica incertidumbre. En este caso, el criterio maximin parecería el más pertinente, ya que todas las alternativas (energía hidráulica, energía fósil, etcétera) tienen aproximadamente las mismas "consecuencias mejores" que la energía nuclear. En la discusión sobre el ADN [51] se ha dicho que una alternativa ("prohibir la recombinación") tiene mucho mejores "consecuencias peores" y mucho peores "consecuencias mejores" que la otra ("permitir la recombinación") así que una combinación sopesada de un racionamiento maximin y maximas [52] sería la apropiada. Sin embargo, en este caso *no hay modo racional de escoger los pesos;* otro ejemplo en que no se define la racionalidad.

Otra posibilidad consistiría en emplear el método de *corazonadas de Gestalt.* Es decir, en lugar de proceder analíticamente y descomponer el problema de decisión en sus subproblemas elementales, a cada uno de los cuales asignamos probabilidades subjetivas del tipo antes mencionado, deberemos operar directamente al nivel del problema global y utilizar nuestro juicio para llegar a una confusión particular. Consideremos el caso de la guerra de Vietnam.[53] Si es verdad que las desastrosas consecuencias de las decisiones

siendo material de armas nucleares, y bien puede decirse con cierta confianza, ¿cuáles son las oportunidades de que un gobierno nacional en un país dado será tentado a valerse de esta opción dentro de cien o quinientos años?

[50] Para las incertidumbres geológicas que hay en el problema de la disposición de los desechos nucleares, *véase,* de Marsily *et al.* (1977).

[51] *Véase* Cohen (1977) para un estudio. Sus conclusiones que se hacen tan sólo para poner de relieve el aspecto conceptual del problema.

[52] *Véase* Luce y Raiffa (1957), pp. 282 *ss.*, 296 *ss.*

[53] Las siguientes observaciones deben ser comparadas con Ellsberg (1971) y Janis (1972).

tomadas por las autoridades militares y políticas se apoyaron notablemente en la teoría bayesiana de la decisión y la asignación de probabilidades subjetivas a subproblemas elementales ("Mi mejor corazonada es que el VC atacará dentro de una semana, con probabilidad de 70 %", etcétera), esto podría tomarse como un experimento natural que refuta el método bayesiano y ofrece un argumento en favor de la superioridad del enfoque *político* global a tales problemas.

No poseo evidencia sólida sobre la verdad del antecedente de la anterior afirmación condicional. Y aun si resultara ser cierto, podría argüirse que esto sólo muestra que los teóricos de la decisión eran malos teóricos y no que habían empleado una mala teoría. Y aun si concediéramos que la teoría es mala, por las razones antes dadas, podría decirse que el enfoque global de los políticos también requeriría probabilidades, corazonadas o conjeturas subjetivas. Mi respuesta a ello sería que la gente sobrevive en política porque logra evaluar las probabilidades subjetivas con el mínimo de distorsión de los procesos causales procedentes a las creencias. (Lo mismo puede decirse para los hombres de negocios que triunfan, lo que no debe confundirse con las empresas que triunfan.) En la medida en que no hay un proceso de selección comparable que separe los que toman decisiones, y que fracasan, en el ejército o en el gobierno, debemos tener extremo cuidado al interpretar las modalidades subjetivas debidas a tales personas. Ahora bien, es un hecho empírico que la gente que sobrevive en política rara vez tiene la preparación formal de teóricos de las decisiones, y sus probabilidades subjetivas a menudo no satisfarán los requerimientos de congruencia normales. Si la información del

proceso de selección política estuviese compuesta tan sólo de teóricos de decisión, entonces el producto de las decisiones podría haber mejorado; pero si las decisiones tuviesen que ser tomadas por teóricos de las decisiones (o sobre la base de evaluaciones dadas por ellos) que no hubiesen estado sometidos a este proceso de discriminación, entonces la calidad del producto sin duda se deterioraría en comparación con el producto dado por gente ordinaria seleccionada por esa cualidad tan fundamental y elusiva: el juicio.

5. Maximización, satisfacción y selección natural

Las últimas observaciones de la sección anterior nos llevan inmediatamente al conjunto de problemas citado en el rubro de esta sección, que habitualmente se asocian a nombres como los de Herbert Simon, Milton Friedman, Gary Becker y Sidney Winter. Repetiré y ampliaré los puntos tratados en la sec. 4, capítulo II, y me explayaré sobre ellos para mostrar que la existencia de una conducta racional bien definida puede ser cuestionada por motivos relacionados con el costo de la información. Para presentar el argumento en el medio apropiado, empezaré por esbozar algunos de los desarrollos de la teoría de la conducta racional que condujeron a ella. En sucesión lógica y parcialmente cronológica, creo yo que podemos distinguir entre cinco posiciones distintas dentro de este desarrollo.

La primera posición es el clásico enfoque de maximización, que presupone total certidumbre o probabilidades objetivas, sin costos de información y de transacción, cálculo instantáneo e

impecable, etcétera. La importancia de este paradigma como norma idealizada siempre deberá tomarse en cuenta. Incluye una perfecta racionalidad, en el sentido del capítulo I, *supra*, así como racionalidad imperfecta, en el sentido del capítulo II. El enfoque está sujeto a las dificultades planteadas en la sec. 2 del presente capítulo, a saber, que el óptimo en algunos casos puede no estar definido, aunque no por las razones que analizamos en esta sección.

La segunda posición es la teoría de la satisfacción o de la racionalidad limitada, que está asociada al nombre de Herbert Simon y cierto número de teóricos posteriores.[54] Los argumentos de esta opinión son en parte empíricos, en parte lógicos. Empíricamente, rara vez se da el caso de que los hombres de negocios busquen deliberadamente el óptimo global en su conjunto factible. Antes bien, tienden a fijarse alguna meta como una tasa mínima de ganancia, y luego adoptar el primer curso de acción que se les ocurre (tal vez como resultado de algún proceso estandarizado de investigación) y que satisface este requerimiento mínimo. Lógicamente podría argüirse que las abstracciones implícitas en el modelo clásico de maximización son demasiado heroicas, y que el modelo nunca podría realizarse, ni siquiera en principio. La búsqueda de lo óptimo cuesta un tiempo que habría podido emplearse con mayor provecho.

La tercera posición toma su clave de la última observación del párrafo anterior. Este es un enfoque modificado de maximización [55] que toma en cuenta, hasta cierto punto, las objeciones de los satisfactores, arguyendo que el satisfacer, debi-

[54] Simon (1954, 1976), March (1977).
[55] Riker y Ordeshook (1973) ejemplifica esta posición.

damente comprendido, es sólo una especie de maximizar con elementos tales como el costo de la información incluido. La posición abstractamente óptima, es decir, la solución a un problema técnico que puede ser óptimo desde un punto de vista de ingeniería puede ser económicamente irracional por ser excesivamente costoso y requerir demasiado tiempo.

La cuarta posición es en cierto sentido, la tercera puesta de cabeza, no la satisfacción como maximización, sino la maximización como satisfacción. La escuela de economistas de Chicago,[56] y Joseph Schumpeter antes que ella,[57] han argüido que la conducta "racional" o maximizadora es, en realidad, resultado de una acción estereotipada o satisfactoria, en el sentido de que los maximizadores son precisamente los satisfactores que son seleccionados por el mercado, o más generalmente por la lucha por la vida, porque casualmente tropezaron con las reglas generales óptimas. Los economistas de Chicago, a partir de Armen Alchian, han presentado varias versiones ligeramente distintas de este argumento. Una que explícitamente niega que las empresas tengan intenciones maximizadoras, y una, según la cual bien podríamos suponer que las empresas tienen otras intenciones porque en todo caso actúan como si las tuvieran.

La quinta posición está representada por un escrito fundamental y un tanto nihilista de Sidney Winter, que inició una larga serie de interesantes obras, muchas de ellas realizadas en conjunto con Richard Nelson.[58] El principal efecto

[56] En particular Alchian (1950), Friedman (1953), Becker (1976).

[57] Schumpeter (1934), p. 80.

[58] Winter (1964, 1971, 1975); Nelson y Winter (1974, 1976).

de este escrito es destruir la tercera posición e implícitamente restaurar la segunda, pero también tiene implicaciones importantes para la cuarta. El sencillo punto de Winter, ya planteado en la sec. 4, cap. II, es que el intento de redefinir la satisfacción como especie de maximización conduce a una regresión infinita, porque la "elección de una estructura de información maximizadora de las ganancias requiere información y no es claro cómo el aspirante a maximizador de las ganancias adquiere esta información, o qué garantiza que no pague por ella un precio excesivo". En cierto punto, esta regresión infinita debe ser interrumpida por la intuición, sin el apoyo de razonamiento formal y, ¿por qué no hacer, entonces, que el punto de corte esté lo más cercano posible de la acción misma?

Winter también tiene convincentes objeciones a las teorías de la conducta racional a partir de la selección natural. Del hecho de que en una situación inicial, una regla general empleada por una empresa determinada es la más lucrativa nunca podremos concluir que seguirá siéndolo cuando tal firma se extienda por relación a otras empresas y modifique así la situación inicial. También muestra que en condiciones verosímiles, maximizadores y no maximizadores pueden coexistir en el equilibrio del mercado, y que los maximizadores pueden extinguirse. Ambas posibilidades son, desde luego, contrarias a la afirmación de que todas las empresas mostrarán una conducta maximizadora de las ganancias porque los no maximizadores son eliminados por la competencia. A estos problemas relacionados con el modelo de selección natural podemos añadir una dificultad esbozada en el capítulo II, a saber, que la selección natural de empresas lo más que po-

dría hacer sería simular una maximización local, a menos que supongamos que todas las posibles reglas generales ya están representadas en el mercado.

El resultado de estas diversas consideraciones parece ser que para el análisis de las decisiones de negocios y probablemente para un gran número de otros problemas de decisión, la teoría de Simon de la satisfacción es la más satisfactoria hasta hoy presentada. También hemos visto que los postulados de satisfacción no pueden derivarse de las suposición de unas intenciones de maximizar la ganancia (como en la tercera posición), ni ser utilizados para derivar una conclusión de conducta maximizadora de ganancia (como en la cuarta posición). Huelga decir que muchos subproblemas serán óptimamente resueltos por los habituales métodos maximizadores, porque "vemos" al punto que el cálculo vale la pena de hacerse, pero yo no creo que los grandes problemas a que se enfrenta la empresa o el individuo puedan resolverse de esta manera. Quiero subrayar que esta conclusión no destruye la tesis planteada en los capítulos I y II. El empleo de estrategias indirectas no presupone una información completa acerca del conjunto factible, y sólo la comparación entre dos o más opciones tiene sus recompensas distribuidas distintamente a lo largo del tiempo.

Las objeciones de Winter a las teorías de la selección natural acerca de la firma no se aplican, en mi opinión, al argumento de selección sobre la supervivencia política esbozado en la sección anterior. Ello ocurre porque los políticos son seleccionados no como resultado de las consecuencias de sus acciones, sino como resultado de las cualidades de juicio que revelan sus accio-

nes. Es difícil conceptualizar esta diferencia, pero un aspecto importante quedaría definido diciendo que "en política nadie se recupera jamás de un solo error de juicio grave". Es decir, la mayor cantidad de capital político puede evaporarse de la noche a la mañana como resultado de un error grave, aun si lo que estaba en cuestión era, esencialmente, de poca importancia. Cuando candidatos a la presidencia pierden la cabeza en accidentes de automóvil o chocan por la impertinencia de un periodista, estos acontecimientos son importantes no en sí mismos sino por lo que revelan acerca del candidato y su probable comportamiento en situaciones que serían mucho más importantes. Por contraste, las firmas no entran en quiebra (aunque puedan despedir gerentes) cuando cometen errores de una índole que podría ser desastrosa en otra situación. Esto significa que la situación de los políticos realmente no tiene nada que ver con la selección *natural*. En la terminología de la sec. 5, capítulo I, es mucho más parecido a un proceso de filtración o de selección *artificial*, en que una conducta determinada se mantiene porque los beneficiarios a la vez reconocen que son beneficiarios y tienen el poder de mantener esa conducta.

6. La conducta tradicional y la conducta al azar

En su obra sobre teoría económica y conducta irracional, Gary Becker distingue, útilmente, entre dos categorías de conducta no racional,[59] es decir, recursos no racionales en el segundo filtro al que antes nos referimos. (Max Weber, inciden

[59] Becker (1976), cap. 8.

talmente, también analizó la conducta tradicional como alternativa a la conducta instrumentalmente racional; sus otros dos casos eran la conducta orientada hacia el valor y la conducta afectiva. Creo yo que la conducta orientada hacia el valor puede analizarse de acuerdo con preferencias lexicográficas, y que la conducta afectiva es similar a la flaqueza de voluntad.) En esta sección, examinaré, independientemente del análisis de Becker el *status* y la importancia de estas dos categorías.

Donde el economista suele explicar la conducta de acuerdo con la elección racional, el sociólogo más a menudo invoca la tradición, los papeles o normas, tratando estos términos como sinónimos para el propósito actual. La diferencia entre los dos enfoques surge muy claramente en el estudio de la delincuencia y el estudio de la educación. En ambos casos, el sociólogo, o alguna caricatura de sociólogo, arguye que la elección de una carrera delictiva o de una educación superior en realidad no es tal elección, sino que el individuo se ve *impelido* a ciertos canales por normas o valores específicos de su subcultura. El economista, en contraste, tiende a suponer que los individuos son *atraídos* por diferentes recompensas relacionadas con los cursos de acción disponibles. Es decir, el sociólogo considera la acción como producto de sus antecedentes causales, y el economista la considera motivada por (la expectativa de) recompensas futuras: causalidad contra intencionalidad. Una implicación verificable de la opinión sociológica es que los cambios del sistema de recompensas debieran tener poca eficacia para producir cambios de conducta, mientras que el economista da por sentado que podemos hacer que la gente se comporte virtualmente de cualquier

manera, haciendo los cambios apropiados al sistema de recompensas.

Más específicamente, William Cobb [60] y Gordon Tullock,[61] entre otros, han distinguido entre dos hipótesis que permiten explicar por qué algunas personas se vuelven delincuentes: la "hipótesis económica" que presupone la maximización de la utilidad esperada y la "hipótesis sociológica" —Cobb la llama la "hipótesis de enfermedad"— que relaciona las actividades delictuosas con normas internalizadas y subculturas delincuentes. Según Tullock, no cabe duda de cuál embona mejor con los datos: también afirma que muchos de los sociólogos que empezaron a probar que el castigo no disuade del delito, terminaron descubriendo lo contrario. Raymond Boudon ha argüido, de acuerdo con lineamientos similares, con respecto a la elección de la carrera educativa.[62] Mientras que escritores como Pierre Bourdieu y Jean-Claude Passeron han considerado que la elección diferencial de una carrera educativa es un mecanismo cultural para la reproducción de las desigualdades de clase,[63] Boudon arguye convincentemente que en gran medida los jóvenes de clase obrera evitan la educación superior simplemente porque es demasiado costosa.

La hipótesis sociológica, en su forma burda, antes esbozada, ciertamente puede ser invalidada por prueba empírica. Además, podemos aducir un simple argumento conceptual y un más importante argumento teórico. El argumento conceptual se origina con Becker y afirma que no es

[60] Cobb (1973).

[61] Tullock (1974).

[62] Boudon (1973a). La edición inglesa de esta obra (Boudon 1974) tiene un análisis menos extenso de esta cuestión particular.

[63] Bourdieu y Passeron (1964, 1970).

posible mantenerse dentro de la tradición si el medio cambia de tal modo que la conducta tradicional queda fuera del conjunto factible.[64] El argumento teórico equivale a una distinción entre dos modos en que los valores pueden influir sobre la conducta. La teoría sociológica simplista, que tal vez nadie ha sostenido explícitamente, dice que los valores determinan *directamente* la conducta, eligiendo un miembro del conjunto factible. Una teoría mucho más verosímil es que los valores determinan las preferencias y que las preferencias y el conjunto factible determinan, en conjunto, la conducta. Esto, creo yo, puede utilizarse para una afirmación más útil de la diferencia entre el enfoque económico y el sociológico. El sociólogo supone típicamente que hay muchas diferencias determinadas socialmente, en las preferencias entre personas; por ejemplo, que los obreros y las mujeres utilizan más tasas altas de descuento al evaluar la utilidad actual de las recompensas futuras (el problema de la "gratificación diferida").[65] El economista, por su parte, tendería a pensar que las preferencias son drásticamente similares, y que la mayor parte de las preferencias de conducta observadas pueden explicarse por medio de diferencias en el control de oportunidades. Otra formulación, no del todo equivalente a la anterior, es que para el sociólogo el método más eficiente, dólar por dólar, de producir, digamos, la igualdad educativa es elaborar el sistema de valores, mientras que para el economista, la manipulación del sistema de recompensas probablemente producirá más resultados. Una tercera manera de plantear la distinción, que no equivale a ninguna de las demás, es que para el sociólogo

[64] Becker (1976), cap. 8.
[65] *Véase* Ainslie (1975) para una revisión general.

el cambio social empieza típicamente con modificaciones del sistema de valores, mientras que para el economista, los cambios del conjunto factible son anteriores e inducen el cambio (endógeno) de preferencias.[66]

Esta reinterpretación tiende a cerrar la brecha entre economistas y sociólogos, en el sentido de que la posición sociológica modificada —y mucho más sensata— reconoce que para cualquier persona dada, la acción es gobernada por preferencias, aun si las preferencias varían con las personas. La brecha también podría —y debería— cerrarse desde el otro lado, haciendo que los economistas reconocieran que hay casos auténticos de conducta estereotipada, tradicional y rígida. Volveré a esta cuestión en la sec. 10 del presente capítulo, y aquí sólo estableceré el punto polémico de que las personas a veces actúan estúpidamente por pura estupidez y rígidamente por pura rigidez. La inercia puede ser un modo racional de enfrentarse a un medio en cambio demasiado rápido, pero también puede ser simplemente lo que es: inercia.

Según la teoría (simplista) de la conducta tradicional, las acciones seguirán siendo las mismas después de un cambio de conjunto factible, suponiendo que la conducta tradicional siga siendo factible. De acuerdo con la teoría de la elección racional, los cambios del conjunto factible a menudo producirán un cambio de conducta. Según el tercer enfoque, que ahora analizaremos, la conducta podría cambiar aun con un conjunto factible constante y sin suposición de preferencias cambiantes. Esta es la teoría de que la conducta a veces es *al azar*, en el sentido de que se realiza de acuerdo con alguna distribución de probabili-

[66] *Cf.* Elster (1976a), para este mecanismo.

dad definida por el conjunto factible. En este sentido, se presupone una conducta aleatoria en muchos modelos de movilidad social,[67] en el modelo de Asch-Cohen para la conformidad [68] y en el modelo de Rapoport, para el Dilema del preso secuencial,[69] para nombrar sólo unos cuantos. En todos estos modelos, las personas son consideradas como portadoras de *estados* y no de *intenciones*, y su conducta como gobernada por antecedentes causales y no por recompensas futuras. La conducta tradicional y la aleatoria corresponden a la imagen causal y no a la intencional del hombre, aun cuando difieran fundamentalmente en otros aspectos.

A menudo es ambigua la interpretación de modelos estocásticos. Son los modelos estocásticos tan sólo aproximaciones útiles a algún mecanismo determinista subyacente (como al arrojar una moneda al aire) o exactamente lo contrario, ¿de modo que presuponemos una causalidad determinista pese a nuestro conocimiento de las perturbaciones aleatorias? En realidad, ambas respuestas pueden ser ciertas simultáneamente. Al nivel más profundo, el arrojar una moneda al aire posiblemente sea gobernado por el proceso cuántico u objetivamente aleatorio; a un nivel más acumulativo (pero aún microscópico), a estos procesos se les puede dar cierta representación determinista muy compleja; y al nivel molecular, esta aproximación misma es acercada por un proceso estocástico el cual, sin embargo, no tiene ninguna relación con el original proceso aleatorio. (La situación es un tanto similar a los casos en que un proceso real en tiempo discreto es representado

[67] Boudon (1973b).
[68] Snell (1965).
[69] Rapoport y Chammah (1965).

matemáticamente en tiempo continuo, lo que con propósitos de computación es representado, luego, en un tiempo discreto que no tiene ninguna relación significativa con el original proceso discreto.)[70] Esto implica que los modelos estocásticos en las ciencias sociales tienen un carácter bastante insatisfactorio; describen, en vez de explicar.

Hay una excepción a esta afirmación: el caso en que el agente actúa deliberadamente al azar. En este caso, las probabilidades de transición (o lo que sea) serían más que una representación compacta de algún determinismo desconocido subyacente en el proceso. Surgirían directamente de la suposición de racionalidad. Un ejemplo típico es el empleo de estrategias mixtas en la teoría de los juegos. Sin embargo, aquí encontramos la afirmación de John Harsanyi, citada antes en la sec. 2 del presente capítulo de que sólo serán elegidas estrategias mixtas equiprobabilísticas por actores racionales y que éstas son generadas por algún mecanismo objetivamente aleatorio dentro de sus cabezas y no por elección deliberada. Esto, si es correcto, implicaría que hemos vuelto, una vez más, a la imagen causal del hombre. Aunque reconociendo que hay poderosos argumentos en favor de la tesis de Harsanyi, yo también puedo dar mis razones para creer que no es adecuada para todos los casos. Sin duda, creo que los esfuerzos de Raymond Boudon, analizados en la sec. 4, capítulo I, por explicar la conducta de movilidad por medio de modelos que incluyen estrategias mixtas son muy dignos de tomarse en cuenta, pero claramente hay una necesidad de más análisis conceptual en este dominio.

[70] Bellman (1961), pp. 67-68; *véase* también pp. 130 *ss.* para unas observaciones breves y sagaces sobre la relación entre modelos determinista y estocástico.

En conclusión, podemos aseverar que los dos casos analizados por Becker pueden ser excepciones auténticas —pero no necesariamente— a la teoría de la elección racional. Sin duda, hay casos en que los valores, las normas y las tradiciones ejercen un efecto directo sobre la conducta, pero las más de las veces esta influencia es mediada por la estructura de preferencia. Y aun si hay casos en que una conducta al azar se debe a una elección deliberada de actuar impredeciblemente, el caso más común es, ciertamente, que la aleatoriedad no es más que la expresión superficial de algún desconocido determinismo causal que opera a espaldas del agente.

7. Explicación del altruismo

En el *corpus* anterior de la bibliografía, los modelos de elección racional a menudo eran asociados a la suposición de que la conducta es motivada tan sólo por propósitos egoístas, edonistas o estrechamente aceptados en sí mismos, y los fenómenos de conducta altruista eran negados o bien se creía que creaban una anomalía insuperable para elegirlos. A un nivel superficial, este eslabón entre la racionalidad y el interés egoísta estrechamente concebido ya no puede sostenerse. Los economistas, de todas las creencias, convienen en que el consumo (o hasta la utilidad) de otras personas pueden entrar como argumento en mi función utilitaria. Mi utilidad puede reducirse como resultado de un aumento del consumo de otras personas (como en la envidia) o puede aumentar (como en el altruismo).[71] Mi bienestar puede de-

[71] De hecho, parece necesaria una definición más fuerte de altruismo, a saber, que alguna reducción en mi propio

pender en parte del consumo de alguna persona o personas específicas (como en el amor)[72] o bien en el nivel medio de consumo (como en la búsqueda de posición social).[73] En la donación altruista, el hecho mismo de dar puede ser positivamente evaluado, por encima de la evaluación positiva del placer que en el donativo encuentre quien lo recibe.[74]

En un nivel más profundo subsisten algunas dificultades. En primer lugar, casi todos los actores parecen creer que sí puede encontrarse una explicación de determinada conducta invocando tan sólo el interés egoísta (estrechamente definido) y esto siempre es preferible, en principio, a las explicaciones que invocan una preocupación por los demás, casi como las explicaciones de elección racional en general son preferidas a los modelos causales de conducta. Como lo detallo más en la sec. 10, *infra*, el economista tiene un momento de triunfo cada vez que logra mostrar la lógica racional de alguna conducta que antes se había considerado irracional, y también creo que tiene la misma sensación cuando muestra que una conducta en apariencia altruista resulta una forma particularmente sutil de interés egoísta. En parte, esta preferencia por la estrecha definición de interés egoísta es puramente metodológica, una preferencia de lo sencillo por encima

consumo sería compensada por aumentos de consumo de los demás para producir el mismo nivel de utilidad. El "altruismo lexicográfico" en que el bienestar de los demás establece una diferencia para mi utilidad sólo cuando mi propio consumo se mantiene constante no sería altruismo, en el sentido en que habitualmente se emplea la palabra.

[72] Becker (1976), pp. 233 *ss.*

[73] Haavelmo (1970); también Hirsch (1976).

[74] Me refiero a la obra inédita de Serge-Cristophe Kolm (París).

de lo complejo. También parece, en parte, basada en una suposición sustantiva acerca de la naturaleza humana: el altruismo no es natural en el hombre. En segundo lugar —y esto va unido a la misma suposición sustantiva acerca de la naturaleza humana— parece existir una difundida creencia de que el interés por los demás, cuando se invoca para explicar la conducta debe, a su vez, ser explicado. Por ello en esta sección analizaré brevemente algunos propuestos mecanismos para explicar la conducta en apariencia altruista como de motivación verdaderamente egoísta, y para explicar el surgimiento de una conducta cuyo carácter altruista *no* sólo es aparente. La reducción lógica o evolutiva de altruismo a egoísmo sigue siendo característica de la teoría de la elección racional.

El paradigma de tales cuestiones es la elección de la estrategia operativa en el Dilema del preso. En términos generales, esta elección puede explicarse siguiendo una de tres distintas líneas de argumento. En primer lugar suponemos que la verdadera matriz de rendimiento es un Dilema del preso, pero que los actores no son racionales. Este enfoque corresponde a la imagen causal y no a la intencional del hombre. Va asociado especialmente a la obra experimental de Anatol Rapoport y sus asociados sobre secuenciales Dilemas del preso, brevemente mencionados en la sec. 6 del presente capítulo. También lo menciona, de paso, Mancur Olson, en el curso de un análisis de los problemas del que aprovecha lo logrado por los demás, en la acción colectiva.[75] En segundo lugar, podemos suponer que los actores son racionales, pero que la verdadera matriz de rendimiento —en contraposición, digamos, con la

[75] Olson (1965), p. 108.

monetaria, no es un Dilema del preso. Más adelante volveré brevemente a esta posición. En tercer lugar, y lo más ambicioso, podemos tratar de mostrar que puede ser racional cooperar aun cuando el juego sea un auténtico Dilema del preso. En la sec. 4, capítulo II, vimos que Descartes arguyó que puede "pagar" ayudar a los demás; aquí comentaré brevemente algunas propuestas recientes.

Creo que podemos desechar, sin análisis prolongado el análisis de "metajuego" ofrecido por Nigel Howard.[76] La asimetría de su "solución" basta, a mi parecer, para descalificarlo como solución en absoluto. Mucho más prometedor es el análisis de "superjuego" ofrecido por Margin Shubik y Michael Taylor.[77] En particular, Taylor ha explorado con gran detalle las condiciones en que la estrategia cooperativa será constantemente elegida por dos jugadores racionales en Dilemas del preso secuenciales, en que las recompensas futuras son descontadas a su valor presente. Sin embargo, yo tengo algunas objeciones a este análisis. Al suponer que las amenazas y las promesas pueden tener aplicación, esto comete petición del mismo principio que se estaba analizando. Además, sólo muestra que (dadas ciertas condiciones) existe un cooperativo punto de equilibrio, y no que hay una solución cooperativa. Por último, y lo de más importancia, las condiciones en que este punto de equilibrio (suponiendo que sea la solución) en realidad se *realizarán*, son tales que hacen probable que el juego subyacente se habrá transformado en algo distinto del Dilema del preso. Las condiciones que aseguran un modo de re-

[76] Howard (1971).
[77] Shubik (1970), M. Taylor (1976a).

solver el dilema también acaban con el dilema mismo. Más adelante volveré a este punto.

Una tercera variedad del enfoque de "racionalidad del altruismo" es la teoría propuesta por Gary Becker en obra reciente.[78] Su análisis no se enfrenta ni puede enfrentarse a situaciones del tipo del Dilema del preso sino que, antes bien, se remite a casos en que un auténtico altruista puede llevar a otras personas (egoístas) a actuar para con él como si también ellas fuesen altruistas, porque pueden obtener más a cambio de lo que dan. Mi objeción a esta ingeniosa teoría es mi sensación de que el auténtico altruista dejaría de ser altruista cuando comprendiera, si lo hiciera, que los falsos altruistas en realidad eran egoístas. El altruismo puede ser puro y desinteresado, en el sentido de que se deriva una utilidad positiva del bienestar de otro, cualesquiera que sean su carácter o su conducta, pero más frecuentemente, se actúa altruistamente hacia alguien como función de su carácter, siendo condición mínima que no esté tratando de aprovecharse de nuestro altruismo (pero *véase* la sec. 3, capítulo IV, para unas excepciones a esta afirmación). La estrategia egoísta de simular altruismo sólo triunfaría si lográramos tenerse oculta, a menos, desde luego, que portarse como si se fuera un altruista terminara por convertir a la persona en altruista. Esta última posibilidad hasta puede ser un incentivo para empezar a actuar altruistamente, en la analogía con el argumento expuesto en la sec 3, capítulo II, y sujeta a las mismas condiciones.

Volvamos ahora del reduccionismo lógico al evolutivo: ¿cuáles son las condiciones que probablemente harán que surja un auténtico altruismo? La biología evolutiva ofrece varios posibles ca-

[78] Becker (1976), caps. 12, 13.

minos para explorar esta cuestión. Por las razones dadas en la sec. 4, capítulo I, no acepto el intento de Gary Becker de extender su "teoría de la interacción social" de las elecciones humanas a corto plazo a la selección natural a largo plazo. También pasaré por alto la selección de grupos, no porque las condiciones requeridas para su observación sean esencialmente imposibles, como en el caso de la teoría de Becker, sino porque parece generalmente aceptado que implican una coincidencia de circunstancias extremadamente improbable.[79] Más prometedores son los mecanismos de selección de parentesco y de altruismo recíproco. La selección de parentesco [80] puede producir una conducta que sea dañina para el actor si es lo bastante benéfica para algún pariente lo bastante cercano. En un ejemplo ofrecido por Richard Dawkins, "un gene, por salvar, mediante suicidio, a cinco de sus primos, no se volverá más numeroso en la población, pero sí un gene por salvar a cinco hermanos".[81] En realidad, la selección de parentesco puede producir una conducta que es *indiscriminadamente* altruista, es decir, que beneficia a todos a la distancia, si las probabilidades son de que haya desproporcionadamente muchos parientes en la vecindad cercana y si es, o bien demasiado costoso (en términos de material genético que podría emplearse mejor en otra parte)[82] o técnicamente imposible reservar la ayuda exclusivamente a los parientes.

[79] Willson (1975) ofrece un buen análisis de este tema.
[80] Hamilton (1964) es el originador de este concepto.
[81] Dawkins (1976), p. 100.
[82] Williams (1966), p. 206. Trivers (1971) ve esto como una "explicación de último recurso", pero no estoy seguro de que se le deba desdeñar tan rápidamente. De la teoría económica podemos aprender mucho acerca de los costos de internalizar las economías externas, y no veo por qué

Hasta donde yo sé, nadie ha propuesto un modelo de trabajo de cómo la selección de los parientes puede producir una conducta altruista en los seres humanos. Sin embargo, Robert Trivers ha argüido que su teoría del *altruismo recíproco* puede explicar actitudes psicológicas como la culpa, la amistad o la gratitud como medios (en última instancia, egoístas) para mantener una conducta altruista.[83] En términos generales, el mecanismo del altruismo recíproco favorece el surgimiento o, como se explicó en la sec. 4, capítulo I, al menos la no desaparición de la estrategia condicional a menudo llamada "toma y daca": "tú me rascas la espalda, y yo seguiré con la tuya".[84] Si a esta estrategia condicional añadimos que la estrategia inicial contra un individuo desconocido siempre es cooperar, resulta claro que en una población de altruistas recíprocos un egoísta mutante se encontraría en desventaja. Recibiría algunos favores de los demás, pero cesarían en cuanto vieran que él no correspondía. Un "sutil engañador" que correspondiera hasta cierto grado, dando menos de lo que recibiera, podría hacerlo mejor, al menos antes de que los demás desarrollaran las contraadaptaciones que les capacitaran a detectar ese "altruismo calculado". Trivers sugiere que la culpa puede ser favorecida por la selección natural "para motivar al engañador a compensar su engaño y a comportarse recíprocamente en el futuro".[85]

Yo carezco de competencia para evaluar lo sólido de este análisis. Sin embargo, debemos obser-

este problema es menos importante en la selección natural.

[83] Trivers (1971).

[84] Dawkins (1976), cap. 10.

[85] Trivers (1971), p. 50.

var que la explicación sólo es reduccionista en el sentido de que se explica la conducta altruista como aquella que, básicamente, es de interés egoísta evolutivo para el organismo, pero no en el sentido que el interés y la preocupación por los demás es simulada, como en el análisis Becker. Por lo contrario, Trivers arguye que la selección natural hará evolucionar la capacidad de detectar los intentos de simular altruismo. El altruismo es más eficiente porque *no* se deriva de un interés egoísta calculado. Otro rasgo importante de la explicación es que no pretende explicar ejemplos específicos de conducta altruista, como, por ejemplo, la tendencia a salvar a una persona que está ahogándose. Los intentos de rescate son explicados por una tendencia general a realizar actos de altruismo, y luego esta tendencia se convierte en objeto de la explicación evolucionista. Como ya se dijo en la sec. 3, capítulo I, este reduccionismo es más satisfactorio que los intentos por reducir la conducta *directamente* a la selección natural.

El surgimiento del altruismo también pudo explicarse por lineamientos estrictamente sociológicos, dado un Dilema del preso como la formación de sindicatos,[86] puede decirse que una interacción prolongada entre los obreros hará que la verdadera estructura de rendimiento diverja de la puramente reglamentaria, porque el bienestar de otros entrará en un Juego del seguro como se definió en la sec. 4, capítulo I, en que la peor alternativa sigue siendo la de ser el incauto, pero la mejor alternativa es *no* ser un aprovechado. El Juego del seguro no tiene una estrategia dominante, de modo que la solución cooperativa sólo

[86] Olson (1965), también Elster (1978a), cap. 5.

se realizará si hay información perfecta. Ahora bien, yo supongo que las condiciones mismas que tienden a transformar el juego, de un Dilema del preso en un Juego de seguros, o sea que los actores forman un grupo pequeño o estable, también tenderá a generar la información requerida para la solución del último juego: el afecto a otros y la información acerca de ellos tienden a crecer *pari passu*. Esto explica la tercera objeción ofrecida antes a la teoría de Michael Taylor. Subraya muy enérgicamente que la conducta cooperativa en el superjuego del Dilema del preso sólo se realizará en comunidades pequeñas y estables, donde probablemente se satisfarán los fuertes requerimientos de información,[87] pero yo acabo de argüir que la interacción sostenida, en tales comunidades, puede resolver el problema mismo al transformar la naturaleza del juego.

Así pues, el altruismo es una anomalía según la teoría de la elección racional sólo en el sentido de que no parece probable que los intentos de explicar la conducta altruista como resultado de un interés egoísta estrechamente definido puedan aplicarse en todos los casos. Altruismo, confianza y solidaridad son fenómenos auténticos que no pueden disolverse en formas de interés egoísta. Este argumento, tomado junto con los comentarios de la sec. 3 del presente capítulo señala la necesidad de una noción más extensa de racionalidad. *El hombre económico* puede ser definido de acuerdo con sus preferencias continuas y su estrecho interés egoísta, pero el *hombre racional* puede tener preferencias no arquimédicas y ser movido por su interés por los demás.

[87] M. Taylor (1976a), p. 93.

8. La inconstancia

La conducta racional, según la definición común, significa actuar de acuerdo con un conjunto de preferencias completo y transitivo. La conducta de un individuo que es racional en este sentido puede ser predicha por él mismo y por otros, siempre que no cambien su preferencia. Sin embargo, ¿qué diremos acerca de una persona que a cada momento muestra preferencias completas y transitivas pero que nunca se fija en una pauta estable? Una falla de la racionalidad en este respecto puede deberse a dos fuentes. La inconstancia en sí misma puede ser una forma de irracionalidad, como puede serlo la incapacidad de adoptar una actitud estratégica hacia las preferencias endógenamente cambiantes. El individuo perfectamente racional está libre de inconstancia. Una "segunda mejor" racionalidad consiste en ser inconstante *y* en tomar precauciones sobre esta preferencia. En la sec. 6, capítulo II, tratamos extensamente este último problema; aquí, la idea de racionalidad "primera mejor" será el enfoque de la discusión. El problema es *cómo elegir las buenas razones para cambiar o no cambiar de opinión, separándolas de las malas razones.* Creo yo que la cuestión es muy importante y que si la resolvemos ello podría traer consecuencias para muchos campos del esfuerzo humano. Por tomar sólo un ejemplo, la teoría de la elección social sería totalmente transformada si pudiéramos dar razones para pasar por alto las preferencias de las personas que constantemente cambian de opinión, para conformarse (o para diferir de) las preferencias socialmente adaptadas. Plantear este problema es, desde luego, mostrar cuán lejos está de ser resuelto; en realidad, yo no implico que pueda

resolverse a satisfacción de todos. A nivel de una mayor generalidad, nos enfrentamos aquí a la cuestión de la *histéresis moral.*[88] Aun si en principio sólo los factores preferentes nos interesan para el propósito de la explicación causal y la predicción, esto no tiene por qué ser así para una evaluación moral. Aquí, la génesis de las preferencias o de la distribución de la propiedad,[89] puede ser pertinente para una evaluación del valor moral.

Entre las buenas razones para cambiar de opinión debemos incluir la experiencia y la enseñanza. La *experiencia* puede formalizarse como la diferencia entre preferencias *ex ante facto* y *ex post facto.*[90] Un paquete de consumo determinado puede parecer más (o menos) atractivo después del verdadero consumo, que antes. Cuanto mayor sea la parte del espacio de bienes ya explorado, presumiblemente menor será el cambio inducido en preferencia después de cada elección. La *enseñanza* ha sido formalizada mediante la distinción entre utilidad y capital de consumo.[91] La enseñanza, en esta conceptualización, no cambia el disfrute que yo derivo de un tiempo dedicado, digamos, a escuchar música, siendo igual todo lo demás. Sin embargo, sí causa una diferencia para el "todo lo demás" al aumentar el capital de consumo que me permite obtener más disfrute de un determinado tiempo de escuchar la música (este enfoque,

[88] Elster (1978b, 1978c) ofrece un estudio más completo de esta noción.

[89] Nozick (1974) sostiene que la histéresis moral es importante para una teoría de la justa distribución del ingreso. Emplea el término "teorías históricas de la justicia", en oposición a "teorías de resultado final", para esta noción.

[90] *Véase,* en particular, Cyert y De Groot (1975).

[91] Stigler y Becker (1977).

debido a George Stigler y a Gary Becker, puede ser un útil replanteamiento de algunos hechos bien conocidos, pero es difícil seguir a los autores cuando también lo emplean para argüir que las preferencias son esencialmente estables. Al transferir el elemento cambiante en la estructura de preferencia a la categoría de capital de consumo, en realidad pueden decir que lo que queda es esencialmente estable, pero se nos debe permitir la duda del contenido sustancial de su afirmación).

De entre las malas razones, debemos incluir el deseo de novedad y su reverso, la pura inercia; también el puro deseo es distinto, y su converso, el conformismo puro. Si una persona siempre desea estar en Nueva York cuando está en San Francisco, y viceversa, esto debe calificar para el epíteto de "irracional". Si alguien tiene un abrumador deseo de la novedad y cambio, podrá ser movido a "mejorarse hasta morir" por una serie de pequeños cambios, cada uno de los cuales es una mejora en términos de las preferencias actuales.[92] Un puro conformista y un puro no conformista pueden perseguirse uno a otro, sin ningún éxito en sus preferencias, como lo saben casi todos los turistas.[93] En tales casos, el individuo es víctima de fuerzas causales que operan "a sus espaldas" y gobiernan sus preferencias de un modo que él mismo comprende. Creo yo que casi todos los lectores acordarán que en muchos casos las razones especificadas son en realidad, las razones. La treta es el número de contraejemplos que inmediatamente acu-

[92] Von Weiszäcker (1971) muestra las condiciones bajo las que esto puede ocurrir.

[93] Gibbard (1974) ha propuesto un modelo formal de este fenómeno.

den a nuestra mente. Leibniz, por su parte, hizo que el deseo de cambio fuese una característica definidora del hombre: "Yo encuentro que la inquietud es esencial para la felicidad de las criaturas, la que nunca consiste en una perfecta posesión que les harían insensibles y como estúpidas, sino en un progreso continuo e ininterrumpido hacia mayores bienes." [94]

Retornaré brevemente a este punto en el capítulo IV, sec. 3, donde se arguye que en el amor (o en algunas formas de amor) es esencial el deseo de cambio y desarrollo. En otros contextos, la extrema forma de constancia que yo he llamado inercia puede ser un rasgo admirable que vacilaríamos en llamar irracional. (No me estoy refiriendo aquí a alguien que deliberadamente manipule sus preferencias para resistir a las fuerzas endógenas favorables al cambio, sino al rasgo de carácter de la inercia *no planeada*). Don Quijote era irracional porque se negaba a adaptarse a las circunstancias cambiantes, pero todos sabemos de casos en que la estatura de una persona aumenta porque no se ha dejado conmover por los vaivenes de la moda.

Sin embargo, pueden proponerse algunas normas formales de inconstancia irracional. En primer lugar, cada vez que una persona "se mejora hasta morir", definitivamente nos enfrentamos a un caso de irracionalidad. Este caso incluye no sólo las preferencias endógenamente cambiantes, antes mencionadas, sino también las estructuras de preferencia intransitiva.[95] En segundo lugar,

[94] Leibniz (1875-1890), V, p. 175.

[95] Raiffa (1968), p. 78, muestra cómo un individuo con preferencias intransitivas podría morir desangrado si se le obligara a dar dinero por cada eslabón de la cadena cíclica de preferencias.

las preferencias que cambian cíclicamente parecerían una marca inequívoca de irracionalidad. Sin embargo, obsérvese que las preferencias que cambian cíclicamente no deben confundirse con la *conducta* que cambia cíclicamente. Esta última puede deberse a la primera, pero también de unas preferencias incongruentes (analizadas en la próxima sección) o a indiferencia, como en la teoría de Harsanyi de las estrategias mixtas, analizadas en las secciones 2 y 6, *supra*, del presente capítulo. En tercer lugar, podemos querer llamar irracional al "dictador invertido" cuyas preferencias siempre coinciden con las preferencias sociales, no porque pueda dictarlas, sino al revés.[96] (Aquí supongo, desde luego, que la coincidencia no es dictada por algunas metapreferencias adoptadas por el individuo, sino que surge por un mecanismo puramente causal.) Yo no creo que estas normas capten todos o siquiera la mayoría de los casos de inconstancia irracional (de la cual, la constancia irracional puede parecer un caso particular), pero, en todo caso, deben bastar para dar al lector una idea del conjunto de problemas que tengo en mente, bajo este rubro.

9. La paradoja

La inconstancia y la incongruencia son nociones interrelacionadas pero diferentes, como ya explicamos en las secciones 5 y 6 del presente capítulo. En el capítulo IV, me enfrento a algunas formas de actitudes incongruentes que son de particular importancia en los asuntos humanos:

[96] Semejante persona también es llamada "camaleón". Para los problemas conceptuales creados por tales individuos para la teoría del poder, *véase* Goldman (1972).

amor, odio, autoengaño. En otra parte [97] he tratado de elucidar la estructura formal de estas contradicciones, y no repetiré aquí estos análisis. Por consiguiente, me limito a algunas contradicciones que relacionan el problema de la incongruencia con la teoría general de la conducta racional.

Paradoja, contradicción e incongruencia han sido analizadas tradicionalmente por poetas y novelistas, y recientemente, también, por psiquiatras, pero rara vez por los científicos sociales. Una interpretación benévola de este hecho es que los científicos sociales han estado tan conscientes de sus limitaciones que han preferido empezar con las facetas más accesibles, aun si menos importantes, de la conducta humana. Yo temo que en muchos casos deba preferirse una interpretación menos caritativa: que han confundido lo accesible con lo importante. En particular, no puedo dejar de sentir que el entusiasmo con que los economistas han aplicado las herramientas del análisis de consumo al amor, el matrimonio, el cortejo o el suicidio está un tanto mal aplicado. No estoy objetando el uso de métodos formales *per se*, sino las triviales suposiciones de que el amor es una externalidad de la función de utilidad [98] o que el valor de la vida eterna debe ser descontado al valor presente.[99]

Las "contradicciones de la mente" son de muchas variedades, y las más reductibles a técnicas formales no siempre son más interesantes. Con esto quiero decir lo siguiente. En algunos casos,

[97] Elster (1978a), cap. 4.
[98] Becker (1976), p. 234.
[99] Azzi y Ehrenberg (1975). A mi parecer, este artículo es una *reductio ad absurdum* contra "el enfoque económico de la conducta humana".

un individuo sostiene un conjunto de preferencias o de creencias de las cuales puede deducirse fácil y llanamente una contradicción, siendo un caso sencillo que el individuo desea o cree *p* y no *p*, simultáneamente. Aquí, las actitudes contradictorias coexisten pasivamente, y es posible que el individuo abandone una de ellas cuando se le haga ver la incongruencia. Por ejemplo, es bien sabido que la mayoría de las personas corrigen sus preferencias si se les hacen ver ciertas intransitividades.[100] Mucho más interesantes son los casos en que la contradicción se debe a un solo *deseo* (o una sola creencia), de donde se puede deducir, de alguna manera, una contradicción. La palabra "de alguna manera" abarca las confusas nociones de implicación conceptual,[101] implicación pragmática [102] y similares, que aquí no pretendo analizar. El lector verá por sí mismo, por los ejemplos presentados aquí y en el próximo capítulo, el tipo de situación en que estoy pensando. En el contexto actual, el punto decisivo es el siguiente. Cuando nos enfrentamos al proyecto contradictorio de, digamos, "tratar de olvidar", no nos enfrentamos a una actitud sin sentido. Podemos comprenderla, aunque no podamos realizarla. Por contraste, si alguien afirma "*p* y no *p*", no podemos captar lo que tiene en mente. Cuando la contradicción o incongruencia se sigue de un solo proyecto unitario, puede ser cons-

[100] Raiffa (1968), p. 75, informa, sin embargo, que algunos sujetos en un experimento prefirieron conservar sus preferencias aun cuando se les señalaran las intransitividades.

[101] Elster (1978a), caps. 4 y 5, ofrece algunas observaciones acerca de esta noción elusiva.

[102] Para un modelo de mundos posibles de la "lógica de la afirmación", *véase* Gullvag (1977).

titutivo de la personalidad en general, y muy difícil de abandonar.

Leslie Farber ha acuñado la frase "desear lo que no puede desearse" para un subconjunto de estos proyectos contradictorios:

> Yo puedo desear conocimiento, pero no sabiduría; irme a la cama, pero no dormir; comer, pero no sentir hambre; mansedumbre, pero no humildad; escrupulosidad, pero no virtud; autoafirmación o desafío, pero no valor; lujuria, pero no amor; conmiseración, pero no comprensión; felicitaciones, pero no admiración; religión, pero no fe; lectura, pero no entendimiento.[103]

Paul Watzlawick y sus asociados han ofrecido muchos ejemplos de la misma índole general: "Tú debes amarme; Yo quiero que me domines; No seas tan obediente; Creí que podía enseñarte a ser un verdadero hombre; Sé espontáneo, etcétera".[104] En mi propia familia hay una anécdota de un remoto antepasado que instruía así a sus hijos: "Hazlo pronto y con gusto." En la sec. 3 del capítulo II, me referí al análisis hecho por Bernard Williams de la imposibilidad de llevar adelante la decisión de creer, y a la observación de Emily Dickinson acerca de la decisión —igualmente imposible— de olvidar. La tesis general del capítulo II también muestra un sentido en que tales decisiones y órdenes *no* tienen que ser paradójicas o imposibles de cumplir, por ejemplo, mediante el uso de estrategias indirectas, haciendo que la acción genere la convicción, en lugar de al revés. Puede dudarse de que la estrategia de "hacer los movimientos" funcione para

[103] Farber (1976), p. 7.

[104] Watzlawick *et al.* (1967), cap. 6; Watzlawick *et al.* (1974), cap. 6. *Véase,* también, cap. IV, *infra.*

todos los casos aquí citados. Puede funcionar para la creencia; pero no para el olvido; para el altruismo, no para el amor; para la virtud, no para un sentido del humor. Sea como fuere hay una huella de paradoja hasta en las simples aplicaciones de la estrategia, pues el que intenta ser altruista puede encontrar, para su consternación, que no es aceptado como auténtico altruista y el que trata de creer puede terminar en un lugar distinto.

Otro conjunto de paradojas ocurre cuando el *status X* nos da el derecho al *status Z*, a condición de cumplir con la acción *Y*, lo que automáticamente nos descalifica de ser *X*. El paradigma es *Catch 22*, sucintamente descrito así por Paul Watzlawick:

> Cualquiera que esté dispuesto a volar en misiones de combate tendría que estar loco, y estando loco, podría ser dejado en tierra, por razones psiquiátricas. Sólo tiene que pedir serlo. Pero el proceso mismo de pedirlo, de no querer volar en más misiones de combate, es prueba de normalidad e impide que lo dejen a uno en tierra por razones psiquiátricas.[105]

Es fácil pensar en variaciones de este género, tal como el uso de formas especiales que hay que llenar por los individuos tan necesitados que ni siquiera son capaces de llenar una forma, etcétera. En la sec. 9, capítulo II, cité el caso del noble ruso inventado por Derek Parfit: una persona que elabora para sí misma una situación en que no puede hacer la petición sin dejar de ser ella misma, y así, perdiendo todo motivo para ser obedecida. El análisis del suicidio hecho por Les-

[105] Watzlawick (1977), p. 25.

lie Farber corre por lineamientos similares, como cuando arguye que "el desesperado se toma su propia vida para demostrar que no es responsable de tomarse su propia vida",[106] olvidando que después de los hechos, ya no estará allí para aprovechar su frustración. Una intención similar se encuentra en el deseo de estar presente en nuestro propio funeral y de oír la oración fúnebre que revela nuestra esencia "tel qu'en lui-même enfin l'éternité le change".[107]

Una nueva categoría de paradoja es dada, paradójicamente, por el dicho de Groucho Marx: "Yo no soñaría con pertenecer a un club que estuviera dispuesto a aceptarme como uno de sus miembros." (En cierto sentido, esta es una inversión de la paradoja del amo y del esclavo, de la que diremos más en la sec. 2, capítulo IV.) Alguien que esté dispuesto a reconocerme a *mí* debe ser, obviamente, indigno de reconocimiento, y su reconocimiento de mí es igualmente inútil. (En la paradoja del amo y el esclavo, la falta de valor del que reconoce, es decir, el esclavo, es el punto de partida, mientras que la falta de valor del reconocimiento, y por ello del reconocido, es la conclusión. Por tanto, puede decirse con verdad que Marx puso a Hegel de cabeza.) Cualquiera que vea jugar a niños pequeños habrá visto esta paradoja, cuando uno se niega a rechazar un objeto que había solicitado porque calcula que debe valer poco si el otro está dispuesto a darlo. También creo que los escritores a veces son víc-

[106] Farber (1976), p. 80.

[107] Cuando Sartre (1943, *passim*) afirma que la contradicción básica del hombre es el deseo de ser simultáneamente *en-soi* y *pour-soi*, cita este verso de Mallarmé como ilustración; pero el episodio funeral de *Tom Sawyer* también procedería. Para la historia de la noción de "volverse el que se es", *véase* Peyre (1963), p. 294.

timas de esa falacia; si alguien elogia su obra, esto lo descalifica automáticamente de ser un crítico serio. Esta es, desde luego, una forma de *hubris* y no de excesiva modestia, pues si alguien no elogia su obra, esto, también es igualmente descalificador.

Yo sospecho que es bastante vana la ciencia social que niegue o se desentienda de tales fenómenos. También creo que hay una importante lección conceptual de sacar de tales casos, a saber, que el *análisis intencional no presupone actores racionales.* Hasta aquí, he empleado tácitamente "racionalidad" e "intencionalidad" como si fueran sinónimos, práctica que se justificaba en los casos antes analizados. Los ejemplos que aquí presentamos muestran que no siempre es así. Las intenciones contradictorias y paradójicas pueden ser inteligibles en función de un proyecto aun si tal proyecto no es racional. Por consiguiente en un sentido, tales paradojas limitan las envergaduras de los modelos de elección racional, pero en otro sentido abren la posibilidad de dar sentido a la conducta que de otra manera sería considerada como patológica y como sujeto tan sólo de análisis causal. Desde luego, el análisis de las facturas contradictorias también debe basarse en modelos causales para decir lo que el actor *hará* en una situación dada, pues aquí el modelo intencional, que predice *demasiado*, no tiene valor.

10. ¿Y CON ESO QUÉ?

Como enérgicamente arguyó Donald Davidson, hay una general *presunción de racionalidad* en los asuntos humanos. En un artículo reciente, Davidson explica lo que le hizo pasar de la imagen

causal a la imagen intencional (o racional) del hombre. Por entonces estaba poniendo a prueba la teoría de Frank Ramsey sobre la utilidad y la probabilidad subjetiva, una temprana —y según Davidson, superior— versión de la teoría de Von Neumann-Morgenstern sobre la utilidad cardinal. Aunque los sujetos a menudo mostraron preferencias intransitivas en experimentos aislados, una sucesión de pruebas mostró algunas rutas más racionales:

> Se descubrió que con el paso del tiempo, las personas se volvían cada vez más congruentes; las intransitividades eran gradualmente eliminadas. Después de seis sesiones, todos los sujetos estaban cerca de ser perfectamente congruentes. Esto bastó para mostrar que una teoría estática como la de Ramsey no podía —ni siquiera en las condiciones más minuciosamente controladas— producir predicciones precisas: simplemente hacer elecciones (sin ninguna recompensa ni *feedback)* altera las elecciones futuras. También hubo un resultado absolutamente inesperado. Si se combinaban las selecciones de un individuo a lo largo de las pruebas, sobre la suposición de que su preferencia "real" era por la alternativa de una pareja que escogía las más de las veces entonces casi no había incongruencias. Al parecer, desde el principio hubo valores subyacentes y congruentes que se realizaban mejor y mejor en la selección. Me resultó imposible elaborar una teoría que pudiera explicar esto, y abandoné mi carrera de psicólogo experimental.[108]

A partir de este experimento y de consideraciones más generales, Davidson llegó a la conclusión no sólo de que "los datos están abiertos a más de una interpretación", sino de que "si atri-

[108] Davidson (1976), p. 107.

buyéramos legiblemente actitudes y creencias... estaremos comprometidos a descubrir, en la pauta de conducta, creencia y deseo, un alto grado de racionalidad y de congruencia".[109] El punto tal vez pueda mostrarse en los siguientes términos paradójicos: para reunir las pruebas mismas sobre las cuales pudiera negarse la racionalidad en el caso de un individuo dado, habríamos de suponer que es racional si la conducta aparentemente observable puede traducirse a las pruebas. La conducta irracional sólo tiene sentido contra un trasfondo de racionalidad; y la pauta siempre puede ser interpretada para que la irracionalidad aparezca en una parte diferente del sistema de conducta.

Este argumento corresponde muy bien a la que durante largo tiempo fue la práctica común entre los economistas. Han trabajado de acuerdo con lo que se llamó *principio de caridad,* afirmando que siempre debemos suponer, como hipótesis de trabajo, la racionalidad de cualquier acción dada, por muy extraña e inadaptada que pueda parecer a primera vista. El científico social debe estar expuesto a derrochar tiempo e imaginación e inventando explicaciones racionales para la acción que observa, y sólo tras repetidos fracasos deberá considerar, tentativamente, irracional esa acción.

Los éxitos del principio de caridad son muchos. En las secciones 4 y 8, capítulo II, analizamos con cierta extensión las razones de por qué pueda ser racional actuar en forma inflexible y rígida, negándose deliberadamente a adaptarse al medio presente. En el capítulo III, sec. 6, observamos que lo que parece una conducta al azar puede deberse a la deliberada adopción de una estrategia

[109] *Ibid.*, p. 108.

mixta. Asimismo, la gente puede tener metas distintas de los que el observador inicialmente supone. La política mercantilista del proteccionismo era irracional si suponemos que la riqueza absoluta era su objetivo, pero no si remplazamos esa meta por el poder (y la riqueza relativa).[110] La conducta de los plantadores en el Sur [de los Estados Unidos] *ante-bellum* bien puede ser irracional si la juzgamos por las normas de maximizar la ganancia, pero no si la juzgamos por las normas de maximizar la utilidad.[111] (Pero *véase* el capítulo IV, sec. 2, *infra* para una afirmación más matizada.) La gente también puede actuar movida por distribuciones de probabilidad y no por valores promedio. La resistencia al cambio técnico en los países subdesarrollados puede ser una respuesta racional a unos métodos que combinan mayores rendimientos promedio con mayores dispersiones en torno de la media, y no un tradicionalismo irracional.[112] También la perspectiva y el tiempo de los actores puede ser más larga o más breve de lo que inicialmente se creía. La falta de respuesta de los campesinos a los cambios de precios parece racional cuando tomamos en cuenta las fluctuaciones de precios en el pasado que hacen racional formarse unas cautelosas expectativas acerca de los precios futuros, aun si estas expectativas resultan erróneas *ex post facto*. Un terrateniente que se niegue a adoptar técnicas nuevas y superiores puede hacerlo porque teme que el mayor rendimiento permita a los aparceros pagar sus deudas y liberarse del *status* de semiservidumbre que a la larga compensaría las ganancias a corto plazo que les co-

[110] Gerschenkron (1970). p. 65.
[111] Elster (1976b); también (1978a), cap. 6.
[112] Wharton (1971); pero *véase* Romasset (1976).

rresponda.[113] Asimismo, las personas pueden tener preferencias cambiantes, y no preferencias incongruentes. Unas elecciones de consumo incongruentes con el axioma de las preferencias reveladas pueden explicarse sobre la suposición de un cambio de gustos. Por último, la gente puede negarse deliberadamente a adaptarse por causa del "costo psíquico" que esto implica. "La mayor de todas las ganancias de monopolio es una vida tranquila." [114]

Algunos de estos gambitos estiran la caridad hasta el límite. En el capítulo II, sec. 4, me refería a la afirmación de Douglass North, de que las *ideologías* son una respuesta racional al problema del "costo de la información". Estos son, supongo yo, es llevar demasiado lejos la suposición de racionalidad. Douglass North, una vez más en otro contexto, explica el conservadurismo de los campesinos como una "decisión de beneficio de costos: los costos psicológicos y de transacción de instituir el sistema de tres campos contraequilibraron los beneficios significativos que prometía".[115] Simplemente, yo me niego a creer que unos campesinos que viven al nivel de subsistencia sacrificarían una mejora de su nivel de vida en bien de una "vida tranquila". También llega un punto en que las expectativas cautelosas dejan de ser racionales. Más generalmente, y como se observó en la sec. 5 del capítulo I, la manipulación de las perspectivas del tiempo para dar racionalidad a una conducta aparente y motivada es, a menudo, una solución *ad hoc* que hay que utilizar con cautela. Este es, en realidad, el

[113] Bhaduri (1973).
[114] Esta frase de John Hicks es citada por Hirschman (1971), p. 55.
[115] North y Thomas (1973), p. 42.

problema general del principio de caridad, que tantos de sus gambitos implican un razonamiento *ad hoc* peligrosamente cercano a tautologías. Postular costos de información, costos de transacción, costos psíquicos o diferentes perspectivas de tiempo sólo para que la conducta embone en la teoría es una solución inaceptable. Siempre habrá que dar alguna evidencia independiente de estas variables adicionales; las hipótesis *ad hoc* deberán tener algún poder independiente de predicción.

Como será claro por las secciones 2 y 5, *supra*, del presente capítulo yo también creo que hay que dar objeciones más radicales al principio de caridad. Las regresiones infinitas implicadas en los juegos sin solución y en el problema de la cantidad óptima de información impiden que el modelo de la elección racional "despegue" del suelo. Esto no significa que las personas en tales casos no puedan hacer elecciones racionales, tan sólo que no hay una elección que sea *la* racional. La situación no es de "una sola salida",[116] pero esto no significa que todas las salidas sean igualmente buenas. Más precisamente, para convertir la situación de una sola salida, deberíamos añadir algunas suposiciones causales a las puramente intencionales, para lograr predecir cuál de las opciones no descartadas será la que se realice.

[116] Para esta noción, *véase* Latsis (1976), pp. 19 *ss.*

IV. LA IRRACIONALIDAD: LAS CONTRADICCIONES DE LA MENTE

1. Introducción

Intenciones no realizadas, deseos frustrados, deseos insatisfechos, proyectos fracasados: esta es la materia de la vida diaria, de la literatura universal, de las ciencias sociales. Algunos fracasos se deben a obstáculos externos, a falsas creencias excusables, accidentes imprevisibles, a riesgos calculados, a la malévola conducta de los demás. Estos casos van desde lo trivial hasta lo trágico, desde Wodehouse hasta *Edipo rey*, su rasgo común es que el fracaso no implica una falta de racionalidad. Surgen otros fracasos cuando las personas actúan racionalmente sobre suposiciones irracionales acerca de la conducta de los demás, por ejemplo, sobre inexcusables creencias falsas. Tales suposiciones pueden brotar en varias formas. En algunos casos, analizados en la sec. 2, capítulo III, los actores suponen, erróneamente, que son los únicos agentes activos y tratan su medio como si fuera constante, o al menos integrado por agentes menos sagaces que ellos mismos. En otros casos, interviene la racionalización de los deseos, como puede ocurrir cuando la gente actúa basándose en las probabilidades subjetivas analizadas en la sec. 4, capítulo III. En una tercera categoría de casos, el autoengaño se encuentra tras las suposiciones racionales. Esta es una fuente más radical de irracionalidad, ya que el agente se las arregla a la vez para creer y para no creer

en las mismas ideas. Exactamente *cómo* se logra esto es el tema de la sec. 4 del presente capítulo.

Por último, las intenciones pueden no realizarse porque son inherentemente irrealizables. Algunas variedades de ello fueron brevemente analizadas en la sec. 9, capítulo III, como "desear lo que no se puede desear", las dobles ataduras, la paradoja de *Catch 22* y la paradoja de Groucho Marx. En las secciones 2 y 3 del presente capítulo exploro el sentido en que odio y amor son emociones que muestran contradicciones internas que les impiden encontrar una satisfacción duradera. En el análisis de estas estructuras me baso en tres tradiciones distintas, que analizaré brevemente antes de ir al grano.

Dentro de la tradición filosófica, la insistencia en las contradicciones como rasgo de la realidad se remonta a Heráclito, y en años recientes se ha asociado al nombre de Hegel. En otra parte he analizado la teoría hegeliana de las contradicciones, concluyendo que debemos distinguir entre la *Lógica*, en que se niega rotundamente el principio de contradicción, y la *Fenomenología del espíritu*, donde se elabora una posición mucho más aceptable.[1] En la primera obra, Hegel dice, en realidad, que dos creencias contradictorias pueden ser ciertas ambas, pero en la última, la principal suposición es que el espíritu humano a menudo mantiene creencias contradictorias, o creencias aisladas de las que se puede deducir una contradicción. De manera semejante, la *Fenomenología del espíritu* analiza las actitudes caracterizadas por deseos contradictorios, sirviendo de paradigma el deseo de comernos nuestro pastel y conservarlo. Después de Hegel, Jean Paul Sartre ha mostrado una aguda visión de estas intenciones

[1] Elster (1978a), cap. IV.

paradójicas. De hecho, *L'être et le néant* arguye que el proyecto fundamental del hombre es el deseo de ser simultáneamente un objeto y una conciencia, *en-soi* y *pour-soi*, argumento elaborado en cierto número de magistrales análisis y actitudes específicas. Como ya observamos brevemente en la sec. 9, capítulo III, la filosofía de Sartre podría resumirse diciendo que el hombre se ve obsesionado por el deseo de escuchar su propia oración fúnebre, de modo que finalmente pueda *saber* qué *es*, dos verbos que son mutuamente incompatibles.

Sartre es tan novelista y dramaturgo como filósofo; y la *Fenomenología del espíritu* ha sido llamada novela filosófica. En realidad, la segunda tradición se encuentra en las expresiones literarias de las contradicciones del espíritu, especialmente en lo que consideramos como el periodo moderno. Para ver la diferencia entre las actitudes clásica y moderna, tomemos la pregunta de Hermione en *Andromaque*: "Je t'aimais inconstant, qu'aurais-je fait fidèle." Desde luego, la respuesta implícita es que el objeto de su amor hubiese sido constante, el amor de ella habría sido aún mayor. La respuesta moderna, por contraste, es exactamente la opuesta: que la inconstancia en el objeto del amor es condición del amor. Aquí, los nombres que acuden a nuestra mente son los de Stendhal y Proust y por medio de ellos, la actitud se ha difundido hasta llegar a ser virtualmente omnipresente. Aquí, no debe interpretarse "moderno" demasiado estrictamente en sentido cronológico. John Donne, escribiendo antes que Racine, tenía una actitud perfectamente afín a la del siglo XX. En realidad, creo que Donne es el escritor con oído más fino para las contradicciones implícitas en casi todos los proyectos humanos. Tomemos, por ejemplo, su análisis concep-

tual de porqué es contradictorio usar el verbo "amar" en tiempo pasado:

> He is stark mad, who ever says,
> That he hath been in love an hour,
> Yet not that love so soon decays,
> But that it can ten in less space devour;
> Who will believe me, if I swear
> That I have had the plague a year?
> Who would not laught at me, should I say,
> I saw a flask of powder burn a day?[2]

[Está loco de remate quien se atreva a decir
que su amor duró una hora;
Y no es que el amor pronto se marchite
sino que en menos espacio puede devorar otros diez.
¿Quién me creería si jurase
que padecí la peste un año entero?
¿Quién no se reiría de mí si dijera
que vi un frasco de pólvora arder todo un día?]

El argumento es que hay una incongruencia pragmática[3] en decir "yo amé", pues el amor transforma a una persona de tal modo que la hace distinta. (El lector percibirá el nexo con la paradoja de *Catch* 22, de la sec. 9, capítulo III.) El extenso uso que más adelante damos a la poesía de Donne justifica un breve análisis de su especificidad. Las frases paradójicas y retorcidas de Donne no siempre han sido fuente de admiración, y hay cierta verdad en la observación de Samuel Johnson: "Las ideas más heterogéneas son unidas con violencia; la naturaleza y el arte son saqueados para obtener ilustraciones, comparaciones y alusiones;

[2] "The Broken Heart"; *cf.* también "The Paradox": "I cannot say I loved, for who can say/He was killed yesterday?"

[3] Hintikka (1967) y Gullvag (1977) se enfrentan a algunos aspectos de esta noción.

y su sutileza sorprende; pero el lector comúnmente piensa que sus logros son costosos y, pese a la admiración que le profesa en algunos momentos, rara vez agrada."[4] Como caracterización de los mejores poemas de Donne, es definitivamente falsa, pero algunas de sus paradojas más explícitas pueden evocar justificadamente esta reacción. Como las paradojas explícitas a menudo son las más útiles para nuestros propósitos, yo no afirmo que los poemas citados aquí sean, uniformemente, buena poesía, aunque algunos de ellos en verdad lo son. Para las cosas peores de Donne podemos remitirnos a sus *Paradoxes and Problems* o a las elegías, donde actitudes son defendidas con un forzado ingenio que sólo a veces se salva por la simple exuberancia de la actuación del poeta.[5] Donne en su mejor forma es demasiado conmovedoramente complejo para poder resumirlo en una simple fórmula, aunque sea una paradoja. Sin embargo, entre los dos extremos hay muchos poemas de muy alta calidad que sí dependen de simples paradojas conceptuales de la índole que aquí analizamos.

Entre mis otras autoridades se cuenta Stendhal, tanto por su visión de las paradojas del amor como por el modo que trató de superar las paradojas de la sinceridad. Julien Sorel y Lucien Leuwen encarnan *las paradojas de De l'amour; Souvenirs d'égotisme* y *Vie de Henri Brulard* tratan valientemente de refutar las tesis de que la sinceridad no es más que la continuación del autoengaño por otros medios. En Stendhal hay matemáticas así como pasión; un espíritu del siglo XVIII

[4] Citado de Clements (comp.) (1966), p. 107.

[5] Una excepción es la Paradoja 10, sobre ingenio y sabiduría, que se muerden a la cola, de la manera más desarmante; *cf.* también Sanders (1971), p. 28.

aliado a una sensibilidad del XIX. Por extraño que parezca este *rapprochement*, estas son las mismísimas cualidades que, para mí, distinguen a Emily Dickinson, citada también a menudo en estas páginas. Su inicial preparación en ciencias se transparenta muy claramente no sólo en sus metáforas tomadas de la botánica o la geología, sino en la precisión formal de sus versos; formal en el sentido de ser acerca de objetos en general.[6] Los autores aquí mencionados —Donne, Stendhal, Dickinson— son voces extrañas, si las comparamos, por ejemplo, a la robustez de Shakespeare, Balzac o Whitman. Además, son extrañas en una forma exclusivamente definida que les hace apropiados para nuestros propósitos. William Blake, para citar a otro escritor extraño, no nos daría el mismo material. Sus paradojas, por ejemplo, sus *Proverbs of Hell*, a menudo son profundas, pero sólo son paradojas en el sentido débil de ser verdades sorprendentes. "El camino del exceso conduce al palacio de la sabiduría", explicado como "Nunca sabemos lo que es bastante hasta que sepamos lo que es más que suficiente", expresa una visión auténtica, y no una actitud incongruente. (Por supuesto, se le puede transformar a una paradoja más fuerte, observando que sería imposible encontrar la experiencia donde está el límite sin cruzarlo, e inconsistente formarse el proyecto de hacerlo.)

Por último, la tercera tradición está integrada por algunos hilos de la historia de la psicología y la psiquiatría. Desde luego, Freud tiene hoy el amor y el autoengaño en el centro mismo de su obra, pero la difícil coexistencia en sus escritos de modelos causales e intencionales del hombre

[6] Hagenbüchle (1974).

hace difícil reconstruir una teoría coherente de las contradicciones a partir de ello. Algunos intentos recientes por replantear la teoría psicoanalítica en "lenguaje de la acción", para librarse de la cosificada terminología de estructuras causales que extrañamente se comportan como actores intencionales, no parecen enteramente logrados, aunque creo que apuntan en la dirección correcta. A un nivel más superficial, el estudioso del autoengaño podrá aprender mucho de la psicología cognoscitiva, de Festinger a Janis y Mann; "superficial" porque estos estudios en su mayor parte evaden el problema filosófico fundamental de cómo es posible el autoengaño, y a veces están demostrablemente en el error, por causa de esta negligencia. Por último, no será sorprendente para el lector que la obra de Bateson, Watzlawick y otros miembros de la escuela de Palo Alto aparezca aquí en relieve. Su tema no es tanto de personas que se fijan metas contradictorias, cuanto de personas que emiten órdenes contradictorias a otros, que entonces quedan atrapados en una "doble atadura" cuando tratan de cumplirlas. En su obra, como en la mayor parte de los otros ejemplos citados en este capítulo, las contradicciones son de índole pragmática y no de carácter directamente lógico; las presuposiciones implícitas de las instrucciones son incompatibles con su contenido manifiesto.

2. El odio

Take heed of hating me,
Or too much triumph in the victory.
Not that I shall be mine own officer,
And hate with hate again retaliate;

But thou wilt lose the style of conqueror,
If I, thy conquest, perish by thy hate.
Then, lest my being nothing lessen thee,
If thou hate me, take heed of hating me.[7]

[Ten cuidado: no me odies
ni te ufanes demasiado en tu victoria.
No es que quiera ser mi propio alguacil
ni pagar tu odio con el mío.
Pero perderás esos aires, propios de quien conquista,
si yo, el conquistado, perezco a causa de tu odio.
Así, para que mi nada no te disminuya,
si me odias, ten cuidado: no me odies.]

Este verso capta tan perfectamente la paradoja del odio que en realidad no hay nada más requerido por el análisis conceptual. Sin embargo, algunas variaciones sobre el tema pueden ser útiles para mostrar cómo y cuándo es probable que surja esta actitud. Ante todo, podemos observar cómo la noción armoniza con el análisis que hace Hegel de la autoconciencia en el capítulo IV de la *Fenomenología del espíritu.* El principal tema de tal capítulo es el esfuerzo del espíritu por enfrentarse, prácticamente, al mundo exterior, no habiendo logrado en los capítulos anteriores llegar a un entendimiento teórico. El proyecto definidor de la autoconciencia en su etapa inicial de *apetencia (Begierde)* queda descrito en los términos siguientes:

> La conciencia tiene ahora, como autoconciencia, un doble objeto: uno, el objeto inmediato de la certeza sensible y de la percepción, pero que se halla señalado *para ella* con el *carácter de lo negativo,* y el segundo, precisamente ella misma, que es la verdadera *esencia* y que de momento sólo está presente

[7] Donne, "The Prohibition".

> en la contraposición del primero. La autoconciencia se presenta aquí como el movimiento en que esta contraposición se ha superado y en que deviene la igualdad de sí misma consigo misma.[8]

Dominar el mundo exterior mediante la apetencia significa asimilarlo físicamente, comiendo y bebiendo (la gratificación sexual aparece en una etapa posterior de la *Fenomenología*, esencialmente con los mismos rasgos paradójicos).[9] Sin embargo, la autoconciencia descubre que esto significa depender del mundo, en lugar de liberarse de él:

> En esta satisfacción, la autoconciencia pasa por la conciencia de la independencia de su objeto. La apetencia y la certeza de sí misma alcanzada en su satisfacción se hallan condicionadas por el objeto, ya que la satisfacción se ha obtenido mediante la superación de este modo; *para que esta superación sea, tiene que ser este otro.* Por tanto, la autoconciencia no puede superar al objeto mediante su actitud negativa ante él; lejos de ello, lo reproduce, así como reproduce la apetencia.[10]

La frase que he puesto en cursivas vuelve a presentar la paradoja en pocas palabras. Si un modo de conciencia se define mediante la negación de

[8] Hegel (1807), p. 135.
[9] *Ibid.*, pp. 262-266.
[10] *Ibid.*, p. 139. Creo yo que las opiniones de Marx sobre la relación entre capitalismo y feudalismo pueden interpretarse de manera similar. El capitalismo sólo es viable mientras puede expanderse por sectores precapitalistas de la sociedad. Con la completa penetración del capitalismo en todos los sectores llega también una baja de la tasa de ganancia y el ascenso de la clase obrera. Por consiguiente, los capitalistas, "se esfuerzan por evitar toda colisión forzosa con la aristocracia; pero la necesidad histórica y los conservadores los empujan hacia adelante" (*New York Daily Tribune*, 25 de agosto, 1852).

otro objeto, entonces la conciencia depende de tal objeto en su mismo ser. La frase "negación interna" a veces se aplica a esta actitud, que también queda ejemplificada en la *decisión de olvidar*, analizada en la sec. 3, capítulo II. Por "tratar de olvidar" algo, yo lo niego y lo afirmo. La orden "olvídalo" puede fijar una doble atadura de una variedad un tanto viciosa, como cuando a un niño le dice uno de sus padres: "Recuerda que no debes ni siquiera pensar en esta cosa prohibida."

Sin comunismo, el anticomunista de la categoría del Dios que falló perdería el significado mismo de la vida; para el ateo militante, la muerte de Dios, o de todos los creyentes, sería igualmente desastrosa.

Habiendo descubierto, con la autoconciencia cuya Odisea está siguiendo, que la apetencia no puede dar satisfacción duradera, Hegel pasa a una actitud más compleja en que la autoconciencia se relaciona con otra autoconciencia. La tradición a esta etapa se efectúa en un pasaje que sigue inmediatamente al texto antes citado:

> Es, en realidad, un otro que la autoconciencia, la esencia de la apetencia; y gracias a esta experiencia ha devenido para ella misma esta verdad. Pero al mismo tiempo, la autoconciencia es también absolutamente para sí, y lo es solamente mediante la superación del objeto y éste tiene que llegar a ser su satisfacción puesto que es la verdad. Por razón de la independencia del objeto, la autoconciencia sólo puede, por tanto, lograr satisfacción en cuanto que este objeto mismo cumple en él la negación; y tiene que cumplir en sí esta negación de sí mismo, pues el objeto es *en sí* lo negativo, y tiene que ser para otro lo que él es. En cuanto que el objeto es en sí mismo la negación y en la

> negación es al mismo tiempo independiente, es conciencia.[11]

Si la autoconciencia puede hacer que el objeto se *niegue a sí mismo*, entonces puede quedar liberada de la perpetua necesidad de encontrar nuevos objetos que negar; o al menos eso creen. Este objeto que se niega a sí mismo, sólo puede ser otra conciencia, que como *esclava* es la propiedad de la autoconciencia original, el *amo*. Los detalles de la dialéctica del amo y el esclavo son tan co nocidos que no los repetiré aquí, pero me referiré a las exposiciones dadas en otra parte.[12] La idea que debemos retener es simplemente que el amo se fija el contradictorio objetivo de lograr un *reconocimiento unilateral*. El objetivo es contradictorio porque para que este reconocimiento tenga algún valor y reproduzca la "satisfacción" deseada por el amo, debe proceder de un ser que él mismo reconozca, y esto no puede ser un esclavo ni un *instrumentum vocale*. El amo no puede simultáneamente alcanzar la satisfacción que obtiene de dominar al esclavo y, mediante él al mundo externo trabajado por el esclavo, *y* la satisfacción derivada del reconocimiento. Y, dicho más simplemente, no podemos obligar a otro a respetarnos. A la paradoja de "sé espontáneo" analizada por Watzlawick y sus asociadas podemos añadir la paradoja "ser respetuoso" de la esclavitud.

Eugene Genovese ha mostrado que en la esclavitud norteamericana esta paradoja tuvo verdadera importancia. La doble condición del esclavo como ser humano y como objeto mostrado, por

[11] Hegel (1807), p. 139.

[12] Elster (1976b), pp. 260 *ss.*; Elster (1978a), pp. 70 *ss.*, 208 *ss.*

ejemplo, en la ley; si el esclavo tiene la capacidad de cometer un delito, ¿no es un delito maltratarlo a él? ¿Cómo puede alguien ser un ser humano como ofensor, y una cosa como víctima?[13] Asimismo, el amo llegó a depender de sus esclavos, por el mecanismo mismo descrito por Hegel: "Los amos necesitaron desesperadamente la gratitud de sus esclavos para definirse como seres humanos morales. Los esclavos, al negarse, clavaron un puñal en el corazón de sus amos."[14] Como en la frase de Donne, en realidad no es necesario que el esclavo sea "mi propio vengador, / Y odio contra odio tome represalias", pues los amos, en su propio interés, se vieron llevados a evitar una brutalidad excesiva, "para que el ser yo nada no te rebaje".

La progresión hegeliana se repite, con algunas variaciones, en la obra de Sartre. La actitud que hemos descrito como odio aparece en Sartre como sadismo:

> De todos modos el sadismo mismo... encierra el principio de su fracaso... En efecto, es la libertad trascendental de la víctima la que trata de apropiarse. Pero precisamente esta libertad que queda, por principio, fuera de todo alcance... *Y cuanto más se encarniza el sádico en tratar al Otro como instrumento, más se le escapa esta libertad.*[15]

Para Sartre, el intento de *poseer la libertad* del otro es característico de muchas relaciones humanas, ideal que es igualmente contradictorio en la forma de amor que en la forma de odio o de sadismo. Sin embargo, en el sadismo, como que-

[13] Genovese (1974), pp. 25 *ss.*
[14] *Ibid.*, p. 146.
[15] Sartre (1943), pp. 475-476.

da indicado en la frase que puse en cursivas, existe la añadida contradicción de que al tratar a la víctima como cosa, el sádico destruye precisamente lo que estaba tratando de apropiarse. Sin embargo, no encontramos el giro por el cual el amo se convierte en el esclavo de su esclavo. Tampoco se encuentra esta visión en los escritos de Freud sobre el tema. Desde luego, Freud reconoce que el sadismo puede convertirse en su opuesto, el masoquismo. De hecho, y en contradistinción con Sartre, Freud ve todo masoquismo convertido en sadismo.[16] En otros casos, el sadismo se convierte en tortura de sí mismo sin un agente externo. El verbo activo se convierte en el reflexivo, y no en la voz pasiva. Sin embargo, estas "transformaciones" son causales y no intencionales; ocurren, pero no podemos comprenderlas. Por contraste, la dialéctica hegeliana es movida por la conciencia que trata de elaborar las implicaciones de sus sucesivos proyectos, de modo que podamos *comprender* cómo la fuerza se convierte en debilidad o el dominio en dependencia. El lenguaje de Freud, de los deseos que cambian su objeto, aunque de alguna manera siguen siendo los mismos deseos, es sumamente difícil de aceptar, y el lenguaje intencional parece mucho más apropiado para estos fenómenos. También creo yo que el lenguaje intencional es más apropiado que el "lenguaje de la acción" pro-

[16] Freud (1969), X, pp. 219 *ss.*, en particular, p. 220: "Ein ursprünglicher Masochismus, der nicht auf die beschriebene Art aus dem Sadismus enststanden wäre, scheint nicht vorzukommen." Sartre, por contraste, analiza el masoquismo antes que el sadismo, aunque no es seguro que este orden sea causal-cronológico, o lógico, fenomenológico (como en el desarrollo de la *Fenomenología del espíritu*).

puesto por Roy Schafer;[17] podemos formar intenciones contradictorias, pero no efectuar acciones contradictorias. La idea de Schafer, de la acción paradójica, es inadecuada porque no capta la idea de que los modos opuestos de conducta se deben a un *solo proyecto unitario.*

3. El amor

Love, oh love, oh loveless love,
We set our hearts on goalless goals,
With dreamless dreams and schemeless schemes,
We wreck our love-boat on the shoals.

[Amor, oh, amor, oh amor sin amor,
unamos nuestros corazones en metas sin metas,
con sueños sin sueños y planes sin planes,
hundimos la barca de nuestro amor contra los bajos.]

Este poema de W. C. Handy tal vez no sea gran poesía, pero nos revela más acerca de las paradojas del amor que el primer verso del poema de Donne cuyo segundo verso fue citado como epígrafe de la sec. 2 del presente capítulo:

Take heed of loving me,
At least remember, I forbade it thee:
Not that I shall repair my unthrifty waste
Of breath and blood, upon thy sights, and tears,
By being to thee then what thou to me wast;
But, so great joy, our life at once outwears,
Then, lest why love, by my death, frustrate be,
If thou love me, take heed of loving me.

[17] Schafer (1976) está muy cerca de Gilbert Ryle (1949) en su conceptualización, mientras que yo creo que la tradición que se basa en Husserl y fue revitalizada por el reciente pensamiento angloamericano resultará, a la postre, más fructífera. Sin embargo, debe añadirse que la demolición, por Schafer, del tradicional lenguaje freudiano, se ha logrado con claridad y fuerza sin precedentes.

[Ten cuidado, no me ames,
o al menos recuerda que te lo prohibí.
No es que vaya a reparar el pródigo desperdicio
de mi aliento y de mi sangre en tus suspiros y lágrimas,
siendo entonces para ti lo que fuiste para mí.
Alegría tan grande nuestra vida al punto la agota.
Así, para que mi muerte no frustre tu amor,
si me amas, ten cuidado y no me ames.]

Esto, supongo, es bastante débil, a la vez como poesía y como paradoja. Las contradicciones del amor se convierten en material sentimental. Empiezo excluyendo la noción de amor irracional como fundada, de algún modo, en falsas creencias [18] acerca de la persona amada. Como ya se dijo en la sec. 1, del presente capítulo, aquí me interesa el amor como proyecto esencialmente irrealizable. También debo decir que no estoy implicando que el amor necesariamente sea paradójico y contradictorio. En realidad, no trataré de definir la emoción; la mejor manera de reconocerlo es a través de sus síntomas clínicos. Estos síntomas pueden corresponder a un proyecto estable, o a otro inestable e incongruente. Aquí sólo me interesa el último pero permítaseme decir brevemente cómo puede conceptualizarse una relación amorosa satisfactoria y estable. Con este fin subrayaré dos características del amor: la transformación del amante y la necesidad de reciprocidad. En conjunto, ambas implican que en el amor las dos personas son mutua y condicionalmente transformadas, y la metamorfosis de cada una depende de la de la otra. Aquí nos basamos en la teoría de Charles Taylor sobre la res-

[18] G. Taylor (1975-1976), pp. 157 *ss.*, tiene un análisis de esta noción.

ponsabilidad de sí mismo, como se analizó en el capítulo II, sec. 9, *supra*. Si otra persona cambia las condiciones limítrofes de mi vida, en tal forma que me permitan redefinirme a mí mismo de un modo que otra manera habría sido inalcanzable, y mi ego redefinido cambia, asimismo, las condiciones limítrofes de esa persona, entonces estamos atados uno a otro en un estado de amor que ciertamente es precario, como la solución de un juego sin estrategias dominantes, pero no necesariamente evanescente o incongruente. Los "interseguros del espíritu" son como "compases gemelos"; en un extraordinario verso, Donne también nos exhorta a "refrenarnos de temores verdaderos y falsos",[19] como si al decirlo pudiese hacer, lo que en realidad puede.

Es cierto que la reciprocidad en este sentido es el objeto del amor,[20] pero también puede significar la muerte del amor. Ya he tocado esto en la sec. 1 del presente capítulo, comentando un verso de *Andromaque*. El punto es enérgicamente planteado por Stendhal en su descripción de la relación de vaivén entre Julien Sorel y Mathilde de la Mole. La conceptualización más sencilla y tal vez la mejor de este fenómeno es por medio de la paradoja de Groucho Marx en el capítulo III, sec. 9: yo no soñaría en amar a alguien que se rebajara tanto que *me* amara a mí. Yenri Peyre atri-

[19] "The Anniversary": las frases anteriores son de "A Valediction: Forbidding Mourning".

[20] Newton-Smith (1973), p. 126, escribe que "parece ser una verdad necesaria que si A ama a B, A desea ser amado por B"; el concepto de necesidad es uno de una conexión necesaria en el caso general, permitiéndose algunas excepciones. Yo no creo que las relaciones de vaivén analizadas en el texto constituyan excepciones a esta afirmación; antes bien, que el deseo no puede sobrevivir a su realización.

buye a Rousseau "esa extraña noción preproustiana de que poseer a una mujer... también es rebajarla",[21] idea que también encuentra su explicación natural por medio de esta paradoja. El amante en este caso no es poseído por la *hubris* implícita en algunas formas de la paradoja, pues ve la no reciprocidad como señal de superioridad y de algo digno de amor. Sin embargo, como en todas las variedades de la paradoja, la reciprocidad es vista como signo de inferioridad. El amante se esfuerza por ser reconocido por una persona cuyo reconocimiento sólo tiene valor mientras no se entrega, en contradistinción con el amo hegeliano que se esfuerza en ser reconocido por una persona a la que ya se ha negado a reconocer, negándole así el estado mismo que sería digno del reconocimiento. (Este contraste también ha sido notado por Sartre, en términos muy similares.)[22]

El amor correspondido puede explicarse, como ya se sugirió, por una analogía que no debe tomarse demasiado literalmente, con la realización de la solución en un juego sin una estrategia dominante. La paradójica forma de amor analizada en el párrafo anterior se parece más a un juego sin solución, analizado en el capítulo III, sec. 2. Podría concebirse un caso intermedio sobre una analogía con los juegos en que la solución

[21] Peyre (1963), p. 99. En la p. 93 Peyre también cita (en su propia traducción) la siguiente carta de Rousseau a Madame d'Houdetot: "Si sois mía, os pierdo, aunque poseyéndoos, a vos, a quien honro."

[22] Sartre (1943), p. 438: "Ce que le maitre hégélien est pour l'esclave, l'amant veut l'etre pour l'aimé. Mais l'analogie s'arrete ici, car le maitre n'exige, chez Hegel, que latéralement et, pour ainsi dire, implicitement, la liberté de l'esclave, au lieu que l'amant exige *d'abord* la liberté de l'aimé."

existe pero no se realiza, precisamente porque no consiste en estrategias dominantes. En esta categoría de casos, en otras palabras, hay una falla de información que impide surgir a la relación. Esta falla de la información puede deberse a causas externas, caso en el cual no es particularmente interesante, pero también puede deberse a la naturaleza de la propia relación. Así, Lucien Leuwen y Madame de Chasteller son incapaces de declararse su amor mutuo, porque las cualidades mismas que les hacen dignos de amor también les impiden hablar. Cuando Madame de Chasteller responde por primera vez a una carta de Lucien, Stendhal observa que

> "¡Ah! Madame de Chasteller responde", habría dicho el joven de París, un poco más vulgarmente educado que Leuwen. "Su grandeza de alma al fin se decidió. He aquí el primer paso. El resto es cuestión de forma; un mes o dos, según que yo tenga más o menos *savoir-faire*, y ella ideas más o menos exageradas sobre lo que debe ser la defensa de una mujer de virtud.[23]

Lucien, por contraste, sólo percibe el tono severo de la respuesta, y si no sabe leer entre líneas para ver que el hecho mismo de haber respondido es el verdadero mensaje. Por otra parte, Stendhal nos dice una y otra vez que es precisamente esta falta de *savoir-faire* la que hace a Lucien aceptable a ojos de Madame de Chasteller; no puede simultáneamente inspirar amor y ser capaz de explotarlo.

Sartre ve el amor como un intento esencialmente vano por "poseer la libertad".[24] (Una ilustración se encuentra en *The Collector*, de John

[23] Stendhal (1952), I, p. 960.
[24] Sartre (1943), p. 434.

Fowles.) El amante desea la sujeción de su objeto, libremente dada. Esta paradoja se refleja, por ejemplo, en una frase cotidiana como la siguiente: "Si me prometes hacerlo sin que yo te lo pida, yo te prometo no pedírtelo." Claramente, estamos tratando con una variedad de la paradoja "Sé espontáneo", pero con el giro, añadido, que puede plantearse de este modo. En la orden común "Sé espontáneo", la contradicción es entre la forma y el contenido de la orden. El mismo contenido en otra forma sintáctica podría ser perfectamente congruente, por ejemplo, "Me gustaría que él fuera espontáneo". La paradoja del amor, en cambio, es más profunda, pues aquí parte del deseo es que se cumpla *como* obediencia a una orden. Esto es, en todo caso, la afirmación de Sartre:

> Así el amante no desea poseer al amado como se posee una cosa: reclama un tipo especial de apropiación. Quiere poseer una libertad como libertad. Pero, por otra parte, no podría satisfacerse con esta forma eminente de la libertad: el compromiso libre y voluntario. ¿Quién se contentaría con un amor que se entregara, como pura fidelidad a la fe jurada? ¿Quién aceptaría oír decir: "Te amo porque me he comprometido libremente a amarte y no quiero retractarme; te amo por fidelidad a mí mismo"? Así, el amante exige el juramento y se irrita por el juramento. Quiere ser amado por una libertad y reclama que esta libertad como libertad ya no sea libre.[25]

Nuevas paradojas se relacionan con la reacción del *tiempo* en el amor. Donne, en varios poemas, explora la importancia del *crecimiento* para el amor. Ciertamente, podemos imputarle una tesis

[25] *Ibid.*

acerca de este elemento, y posiblemente, también una ligeramente más fuerte. La tesis débil es esta:

Except our loves at this noon stay,
We shall new shadows make the other way,
As the first were made to blind
Others; these which come behind
Will work upon ourselves, and blind our eyes.
If our loves faint, and westwardly decline;
To me thou, falsely, thine,
And I to thee mine actions shall disguise.
The morning shadows wear away,
But these grow longer all the day,
But oh, love's day is short, if love decay.
Love is a growing or full constant light;
And his first minute, after noon, is night.[26]

[Si nuestros amores no permanecen en el cenit,
proyectaremos nuevas sombras en otra dirección.
Así como los primeros estaban destinados a cegar a otros; los que vinieron detrás
influirán en nosotros y nos cegarán.
Si nuestros amores decaen y hacia el poniente declinan,
disfracemos nuestros actos; tú los tuyos falsamente
míos; yo, los míos, falsamente tuyos.
Las sombras matutinas desaparecen,
pero éstas se alargan todo el día.
Qué breve es el día del amor si el amor declina.]

Si la satisfacción *(Befriedigung)* derivada del amor es una función a la vez de la intensidad de la emoción y de su ritmo de cambio, este texto dice que si la segunda variable adopta valores negativos, entonces el valor de la función es cero. Comparémoslo con Emily Dickinson sobre un tema similar:

[26] "A Lecture upon the Shadow".

Crumbling is not an instant's Act
A fundamental pause
Dilapidation's processes
Are organized Decays

Tis first a Cobweb on the Soul
A Cuticle of Dust
A Borer in the Axis
An Elemental Rust —

Ruin is formal — Devil's work
Consecutive and slow —
Fail in an instant, no man did
Slipping — is Crash's law.[27]

[La desintegración no es un acto instantáneo.
Hay una pausa fundamental,
un proceso de dilapidación,
un declinar organizado.

Es, primero, una telaraña en el alma,
una cutícula de polvo,
un taladro axial,
una herrumbre fundamental.

La ruina afecta la esencia —obra del demonio,
consecutiva y lenta.
Nadie se malogró en un instante.
Deslizarse es la ley que rige la caída estrepitosa.]

Ambos pueden tener razón, porque no necesariamente están interesados en los mismos procesos. La *previsión* de la descomposición puede hacer que la descomposición ocurra instantáneamente, que el futuro negro actúe en retrospectiva sobre el presente, como en un Dilema del preso secuencial.[28]

[27] Dickinson (1970), núm. 997.

[28] Cuando el futuro se conoce con precisión, es decir, el número de juegos que se jugarán es finito y conocido

La tesis más fuerte dice que el amor debe crecer si quiere existir; que la función antes definida adopta el valor cero si cualquiera de las variables es cero; en el capítulo III, sec. 8, citamos una afirmación similar de Leibniz;[29] imaginemos también una carrera cuesta abajo en bicicleta o esquí, en que el placer depende a la vez de la velocidad y de la aceleración. El fundamento psicológico de esta tesis es muy diferente que en el anterior; la falta de una experiencia real de desarrollo, y no la previsión de descomposición. No es seguro que pueda imputársele a Donne, quien sólo afirma que es posible un amor que crezca indefinidamente:

> And though each spring do add to love new heat,
> As princes do in action get
> New taxes, and remit them not in peace,
> No winter shall abate the spring's increase.[30]
>
> [Y aunque cada primavera dé nuevo ardor al amor,
> así como a los príncipes durante las hostilidades
> se les otorgan nuevos impuestos, a los que no
> renuncian

de antemano, entonces la racionalidad finita de la estrategia no cooperativa del último juego tiene un efecto contagioso sobre él en el último que, para todos fines prácticos parece ser el último; y así hasta el primer juego de la secuencia. Un programa similar surge en la "paradoja del examen", cuando se dice a cierto número de estudiantes que serán sometidos a un examen inesperado durante la semana próxima; entonces es fácil mostrar que no puede ocurrir ningún examen más. A menudo se ha observado que la vida humana sería completamente transformada si los hombres fueran inmortales; yo creo que una transformación no menos profunda ocurriría si conociéramos de antemano la duración de nuestra vida.

[29] *Cf.* Elster (1978a), p. 64, para algunas observaciones sobre la estructura matemática de esta opinión.

[30] "Love's Growth"; *cf.* también "Lovers' Infiniteness".

llegada la paz, tampoco el invierno abatirá el medro primaveral.

Estos son fenómenos paradójicos, no en el sentido estricto de proyectos inherentemente irrealizables, sino en el sentido menos firme de emociones inherentemente inestables y evanescentes, que sólo se presentan al surgir, o que ya se han ido cuando desaparecen.

El tiempo también sirve para la paradoja por un mecanismo diferente, relacionado con la estructura del tiempo interno. Como ya lo explicó Husserl,[31] la conciencia del tiempo interno es una red de intenciones temporales: retenciones, impresiones y propensiones, dirigidas hacia el pasado, el presente y el futuro, respectivamente. El objeto de cualquiera de tales intenciones es una experiencia integral y completa con sus propias retenciones, impresiones y protensiones. Esto significa que cuando yo capto el pasado, también capto el presente como el futuro de tal pasado; de manera similar, el presente está encarnado en el futuro como lo que entonces habrá sido. Estas correspondencias pueden ser no sólo latente o implícitamente conocidas de la conciencia, sino que siempre contribuyen algo a la tonalidad del presente, haciéndole a la vez, la realización de una expectativa y el nacimiento de un recuerdo. En el amor, estos tonos pueden adoptar un predominio patológico, de modo que la actual reunión con el amado sea experimentada *sólo* como el lugar de reunión de pasadas expectativas y futuras lamentaciones. La experiencia puede ser ricamente cómica en sus manifestaciones exteriores, pero internamente la paradoja se vive con

[31] Husserl (1966), en particular §§ 24, 25.

el mudo frenesí bien descrito por Stendhal en el capítulo XXIV de *De L'amour*:

> Cuando se debe ver por la noche a la mujer que se ama, la espera de tan grande dicha hace insoportable todos los momentos que nos separan de ella.
>
> Una fiebre devoradora nos hace emprender y abandonar 20 ocupaciones. Miramos el reloj a cada momento, y nos sentimos encantados cuando vemos que han podido pasar 10 minutos sin mirarlo; por fin suena la hora tanto deseada, y cuando estamos a su puerta, dispuestos a tocar, no nos preocuparía no encontrarla; sólo por reflexión nos sentiríamos afligidos; en una palabra, la espera de verla produce un efecto desagradable.
>
> Estos son esas cosas que hacen decir a las buenas gentes que el amor hace desvariar.
>
> Es que la imaginación, violentamente retirada de ensueños deliciosos en que cada paso produce la dicha, es obligada a enfrentarse a la severa realidad.
>
> El alma tierna sabe bien que en el combate que va a comenzar en cuanto la veamos, la menor negligencia, la menor falta de atención o de valor será castigada por una derrota que emponzoñará durante largo tiempo los ensueños de la imaginación.

Este desplome del tiempo interno está lejanamente relacionado con otra paradoja del amor, ligada con el tiempo, que también gira en torno de la previsión de futuras lamentaciones. En las situaciones descritas por Stendhal, el amante tiene, a menudo, la certidumbre de que lamentará lo que haga. La estrategia obvia para enfrentarse a este problema consiste en simular una decisión, y luego hacer exactamente lo propuesto, con la esperanza de sorprenderse a sí mismo desprevenido. "Siendo la clase de persona que soy, este es el tipo de error irreparable que cometeré, por lo que

estaré en terreno seguro si cedo al impulso opuesto." Pero entonces, de forma inevitable, viene la segunda reflexión: ¿qué pasa si *esta* estrategia es el tipo de error por el cual todo será destruido? La estructura general de este razonamiento está relacionada con el Problema de Newcomb, paradoja analizada inicialmente por Robert Nozick.[32] Aquí, una persona recibe instrucciones resumidas de esta manera:

> En la mesa, ante nosotros, están dos cajas, A y B. La caja B, contiene 100 dólares, la Caja A contiene cero dólares o bien 1 000 dólares. Tenemos la elección, entre abrir A y tomar, como recompensa, lo que tiene dentro, o abrir A y B y llevarnos como recompensa lo que está dentro de ambas. El dinero fue colocado dentro de la caja A *antes* de que hagamos la elección, por un Ser que en el pasado ha tenido una serie ininterrumpida de un millón de adivinanzas atinadas sobre qué elección harán los otros sujetos (dadas estas mismas instrucciones) después de colocar el dinero, y congruentemente ha colocado el dinero de modo que todo el que abra A ha recibido 1 000 dólares y todo el que ha abierto ambas cajas sólo ha recibido 100 dólares. En otras palabras, cuando el Ser prevé que alguien abrirá ambas cajas, no coloca nada dentro de la caja A. De otra manera, coloca en ella 1 000 dólares, y siempre ha atinado hasta hoy en sus predicciones. Ahora, haced vuestra elección.

El dilema de la persona que recibe estas instrucciones es obvio. Su mejor opción parece tratar de ser más sagaz que ese Ser; analizarse a sí misma para ver si es el tipo de persona que probablemente sólo abrirá la caja A y, en tal caso, abrir ambas cajas. El truco, desde luego, está en que no puede estar seguro de que no es, en cambio, el

[32] Nozick (1969b); también Elster (1978a), pp. 85 *ss.*

tipo de persona que probablemente calcularía de esta manera, caso en el cual el Ser no colocará ningún dinero en la caja A. Vemos que ésta es en realidad una variedad del intento paradójico de ser simultáneamente *en-soi* y *pour-soi*. El amante o el que recibe las instrucciones trata de contemplarse a sí mismo desde el exterior, como si fuese una esencia fija por una naturaleza inmutable, y luego añade un grano de conciencia para explotar este conocimiento así adquirido. Esto es como un intento de sorprender nuestra propia sombra dándonos vuelta rápida e inesperadamente. El amante bien siente que *hay* una acción que puede salvarlo, a saber, la acción que no elegirá: de aquí es necesariamente llevado al intento de ser un *pour-soi* contemplando su *en-soi*. Sartre cree que este proyecto es constitutivo de la realidad humana, pero tal es una conclusión extrema e infundada. El deseo de sorprenderse a sí mismo cuando no está prevenido surge de la compulsión de situaciones específicas, ya sean planeadas por un ser omnisciente o por un amor imposible.

4. El autoengaño

> Il y a donc deux manières de falsifier: l'une par le travail d'*embellir*, l'autre par l'application à *faire vrai*. [Hay, pues, dos maneras de falsificar: una de ellas por el trabajo de *embellecer*, la otra por la aplicación a que *parezca verdad*.] [33]

Aquí Valéry, adelantándose a Sartre,[34] dice que el autoengaño puede alojarse en la sinceridad, así

[33] Valéry (1957), p. 571. *Cf.* p. 570: "C'est une loi de la nature que l'on ne se défende d'une affectation que par une autre."

[34] Sartre (1943), pp. 102 *ss.*

como en la insinceridad. El hombre simplemente no puede dejar de ser falso ante sí mismo, y la orden "Sé auténtico ante ti mismo", es imposible de cumplir. La omnipresencia y la necesidad del autoengaño es una tesis extrema. En el otro extremo se encuentra la afirmación de que el autoengaño es imposible, y esto entrañaría que la misma persona a la vez conocería y no conocería la misma cosa y al mismo respecto. Entre ambos casos se encuentra el sentido común, que nos dice que los hombres a veces pero no siempre se engañan. Trataré de mostrar que esta posición está en armonía con la teoría de la racionalidad imperfecta, esbozada en el capítulo II, aunque no afirmo que este enfoque pueda encargarse de todas las variedades del autoengaño. Algunos de los casos restantes pueden ser tocados con adicional maquinaria conceptual, y otros casos más quedarán sin solución. Este éxito parcial no es una gran realización, ya que puede decirse que los casos que quedan en suspenso son los difíciles. Por otra parte, no se puede decir que ninguna otra teoría ha resuelto completamente este problema, que, a mi parecer, es la prueba fundamental que tiene que pasar toda teoría de la naturaleza humana. La teoría freudiana de la mente es notoriamente incapaz de dar sentido a la idea de autoengaño en resistencia y defensa,[35] pese a recientes refinamientos.[36] Sartre, así como en otras partes, re-

[35] *Ibid.*, pp. 88 *ss.*, Gardiner (1969-1970), Schafer (1976), pp. 235 *ss.*; también Fingarette (1969), cap. VI, y Pears (1974).

[36] Schafer (1976), mientras critica el tradicional marco freudiano, trata de reafirmar y de resolver el problema del autoengaño en su "lenguaje de la acción". Sin embargo, yo no considero que los conceptos básicos estén suficientemente aclarados para que esta reinterpretación sea convincente. Para una interpretación típica —y para mí

suelve el problema por el método de tener más fuerza que el enemigo, haciendo que el autoengaño sea tan omnipresente que desaparece como fenómeno distinguido. Fingarette, en el más reciente análisis completo,[37] tropieza con el punto donde hay que explicar el surgimiento de un ego unitario. Los recientes avances de la psicología cognoscitiva [38] no parecen haber captado siquiera la naturaleza de la dificultad. El autoengaño puede ser, en realidad, una forma de "cognición caliente",[39] pero también tiene rasgos específicos que exigen un enfoque diferente.

Permítaseme empezar con cierto número de observaciones clasificatorias, para colocar el autoengaño en el mapa de las contradicciones del

incomprensible— considérese lo siguiente: "...la gente frecuentemente resuelve en forma consciente, preconsciente o inconsciente pensar sólo en ciertos pensamientos, sentir sólo ciertas emociones y efectuar sólo ciertos actos" (p. 197). Cómo lograr semejante resolución inconsciente es para mí un total misterio; un replanteamiento del problema, no una solución.

[37] Fingarette (1969), pp. 107 *ss.* En la p. 109 afirma que "el mundo es tal que puede hacer que los individuos se vuelvan personas"; es decir, que la comunidad de egos adquiera un miembro dominante e integrante, quedando los otros subordinados y transformados por este ego central. Añade más adelante que cómo ocurre esto "es cuestión que no necesitamos debatir aquí". Yo creo, por lo contrario, que su modelo del ego y de la conciencia requiere una explicación de este proceso si quiere que se le tome en serio. Algunas de las ideas del capítulo II, *supra*, pueden ser pertinentes aquí, si el surgimiento de un ego entre los egos aparece como resultado de planeación estratégica (y posiblemente, de contraplaneación). Sin embargo, en este caso el problema del autoengaño puede tratarse de manera totalmente distinta, como decimos más adelante.

[38] *Véase* en particular la reciente síntesis de Janis y Mann (1977).

[39] Abelson (1963) se encuentra en el origen de esta noción.

espíritu. Para empezar, como lo han observado muchos escritores,[40] el autoengaño está relacionado con la flaqueza de voluntad, ya sea como analogía cognoscitiva con tal fenómeno o como condición de posibilidad. Al engañarse a sí mismo en cuanto a la naturaleza del bien, el hombre puede hacer el mal mientras sabe "en su corazón" lo que es bueno. Claro es que ésta no puede ser una solución al problema de la acrasia. Para empezar, existen los casos comunes de flaqueza de voluntad en que yo hago el mal con plena conciencia del bien. Y en segundo lugar, la explicación, cuando conviene, sólo es un caso de *obscurium per obscurius*. La analogía formal entre los dos fenómenos es más interesante que los supuestos nexos causales. En el nivel más general, ha habido intensas discusiones sobre si el autoengaño y la flaqueza de voluntad son siquiera posibles; si representan nociones contradictorias acerca de fenómenos mentales, y no nociones acerca de fenómenos mentales contradictorios.[41] En un nivel más específico, ambas nociones incluyen un conflicto en que hay razones de ambos lados, y el lado más fuerte pierde ante el más débil. Este es, obviamente, el caso de la flaqueza de voluntad. Para ver que también se sostiene para el autoengaño, consideremos el caso de una persona que deliberadamente se niega a reunir todas las pruebas disponibles porque teme que vayan contra algunas de sus creencias más caras. La falta de esta prueba cuenta como prueba de una índole para su fe, aunque en realidad es una prueba dé-

[40] Rorty (1972), p. 405; Fingarette (1969), pp. 76 *ss.*

[41] Para esta distinción, *véase* Elster (1978a), pp. 67 *ss.* También Rorty (1972), p. 395: "Las atribuciones de autoengaño son atribuciones de incoherencia, no atribuciones incoherentes."

bil. Más adelante volveré a este punto. Aquí, tan sólo señalaré algunas diferencias entre acrasia y autoengaño. Además de la diferencia obvia entre las esferas práctica y la teórica en que estas paradojas surgen, existe la desemejanza fundamental de que el autoengaño, en contraste con la flaqueza de voluntad, es un *proyecto* intencional. Podemos, al menos en términos generales, conceptualizar la flaqueza de voluntad, como cierta clase de *sordera* mental [42] en que los proyectos intencionales de la mente son sofocados por mecanismos causales. El autoengaño, en cambio, *es* esencialmente un proyecto, el proyecto de engañarse a sí mismo; surge la paradoja porque "je dois savoir tres précisément cette vérité *pour* me la cacher soigneusement".[43] [Debo conocer precisamente esta verdad *para* ocultármela cuidadosamente].

Como queda implícito en esta última observación, y correctamente subrayado por Fingarette,[44] el autoengaño debe ser claramente distinguido del mantenimiento simultáneo de creencias incompatibles. Implica esto, pero también algo más, a saber, el elemento intencional al que acabamos de referirnos. También debe distinguirse el autoengaño de la racionalización de los casos, donde no hay ni contradicción ni intencionalidad.[45] La racionalización de los deseos surge porque somos causalmente influidos por nuestros deseos y preferencias al evaluar la evidencia. En cierto sentido, está más cerca de la flaqueza de voluntad que

[42] Davidson (1969), p. 113.
[43] Sartre (1943), p. 87; las cursivas son suyas.
[44] Fingarette (1969), cap. II.
[45] Para unos análisis de la relación entre autoengaño y la racionalización de los deseos, *véase* Fingarette (1969), páginas 19-20, y Pears (1974).

del autoengaño. Los procesos psíquicos que ocurren, sin que yo lo sepa, pueden inducirme a seguir las razones más débiles por encima de las razones que conocemos con más fuerza (flaqueza de voluntad) o hacer que las razones más débiles parezcan más fuertes (racionalización de los deseos); ambos casos son diferentes del intento deliberado de pasar por alto las razones más poderosas. Los estudios de la toma de decisiones hechos por la psicología cognoscitiva enfocan intensivamente la racionalización de los deseos, bajo el término de "evitación defensiva". En el reciente y exhaustivo estudio hecho por Janis y Mann, muchas de las subestrategias de evitación defensiva parecen muy cercanas al autoengaño, como cuando los sujetos se convencen a sí mismos de pasar por alto las consecuencias remotas de sus acciones.[46] (Aquí, desde luego, los casos también pueden ir acabando como flaqueza de voluntad.) De modo sorprendente, en su trabajo no hay ningún análisis del autoengaño y ninguna referencia a un cuerpo de bibliografía filosófica en este campo. A este respecto perpetúan la ingenuidad filosófica de otras anteriores de psicología cognoscitiva, en particular de la escuela de Festinger. Las consecuencias de este descuido se analizan más adelante.

También hay que quitar algunos obstáculos relacionados con la sinceridad. Como lo ha mostrado Henri Peyre en su extenso estudio de la sinceridad en la literatura,[47] muchos autores han asimilado la sinceridad con la incongruencia. Presentarse uno mismo como hecho de una sola pieza, es necesariamente un engaño, de sí mismo o de los demás; la auténtica sinceridad implica

[46] Janis y Mann (1977), p. 92.
[47] Peyre (1963), pp. 38 *ss.*, 88 *ss.*, 319 *ss.*

que todas las facetas contradictorias del ego reciben libre expresión. Esto contrasta marcadamente con el enfoque de Fingarette, en que un tema principal es que el autoengaño surge porque el ego no ha alcanzado aún una estructura integrada y en cambio mantiene vivos, tras bambalinas, algunos de sus proyectos oficialmente rechazados.[48] A cierto nivel es fácil ponerse del lado de Fingarette y de Peyre, cuando este último observa:

> Moral y psicológicamente, la aceptación pasiva de nuestras contradicciones y su exposición al analizar ciertos personajes ficticios que observan todo síntoma de desunión en ellos mismos y registran todo lo que es innoble en sus deseos o en sus sueños puede no ser muy superior a establecer una jerarquía, un orden de prioridad entre todos los afanes que solicitan nuestra atención.[49]

La *hubris* de confesión, "profesar culpa para darse crédito por el pecado", como, según fama, dijo John von Neumann de Robert Oppenheimer,[50] es en realidad un tema constante de la bibliografía sobre la sinceridad. Tanto Valéry como Sartre acusan de esta falla a Stendhal,[51] aunque yo no puedo dejar de sospechar que éste es un caso de autopromoción. Sea como fuere, debemos añadir que también es posible ser más o menos honrado hacia nuestros propios autoengaños. El autoengañador reincidente que se niega a cobrar conciencia de su hábito doblemente se engaña a sí mismo, mientras que el autoengañador que conoce

[48] Fingarette (1969), cap. IV.
[49] Peyre (1963), p. 322.
[50] Ulam (1976), p. 224.
[51] Valéry (1957), p. 570; Sartre (1943), p. 105. Contra esto, yo prefiero la interpretación de Wood (1971), p. 100.

su propensión a este respecto puede enfrentarse estratégicamente a su problema, como ya se explicó en el capítulo II, *supra*. La unidad de la mente no implica ausencia de contradicciones, tan sólo una conciencia de ellas y la capacidad de reducir sus defectos sobre la conducta.

La clave a mi propio enfoque al autoengaño es la noción de *responsabilidad para consigo mismo*, con las variedades aristotélica y sartreana distinguidas en el capítulo II, sec. 9, *supra*. Muchos aparentes casos de autoengaño pueden ser reinterpretados como intentos de automodificación (triunfales o frustrados). A cada una de las dos subvariedades corresponde una forma particular de autoengaño, como paso a explicarlo.

El autoengaño asociado a la formación de carácter, en el sentido aristotélico, ya ha sido analizado con cierta extensión en el capítulo II, sec. 3. Casos paradigmáticos son la decisión de creer y la decisión de olvidar. Éstos, si se llevan adelante, ciertamente calificarían como ejemplos de autoengaño. Tales decisiones pueden presentarse en tres modalidades. En un extremo está el intento de llevar adelante, digamos, la decisión de creer de manera directa y plenamente consciente; esto, como lo ha argüido B. A. O. Williams, es conceptualmente imposible. En el otro extremo se encuentra el uso de estrategias indirectas, con el doble efecto de producir el estado deseado de fe y de producir el olvido. (Una estrategia similar es analizada por Fingarette, en un ejemplo sobre un hombre que se emborrachó hasta caer en el estupor, para poder autoengañarse.)[52] En este caso, la persona en realidad está engañándose "a sí misma" pero la decepción sólo puede tener éxito si la anterior y engañadora encarnación del

[52] Fingarette (1969), pp. 31, 80.

ego puede preparar el escenario de modo que pueda ser muerta por el ego posterior y engañado. Una falta de continuidad del yo es la condición para el buen autoengaño en este caso, así como la falta de integridad puede ser una condición en otros casos. Entre estos "otros casos" —los "casos difíciles" referidos antes— se encuentra la tercera modalidad de decidir creer, intermedia entre las dos variedades antes definidas. En este modo, el autoengañador intencional e instantáneamente adopta una nueva creencia pero de alguna manera logra ocultarse esta intención a sí mismo.[53] ¿Cómo lograr esto? Tal se ha convertido en *el* problema del autoengaño. No propondré una solución, salvo para expresar la esperanza de que la clase residual de casos que no pueden recibir un análisis más directo acabará por reducirse a un cascarón vacío.

Ya he citado el verso de Donne "Refrenémonos de temores verdaderos y falsos". La exhortación de evitar los falsos temores es bastante convencional, pero la referencia a temores verdaderos se parece extrañamente a una invitación al autoengaño. Sin embargo, como ya sugerimos antes, se le debe comprender, mejor, como una invitación a la automodificación en el sentido sartreano del término. Este enfoque ha sido dirigido, en realidad, por varios recientes escritores y estudiosos del autoengaño.[54] Aun si no siguen a Sartre en su hincapié en la irrestricta libertad del ego para definirse a sí mismo ("il n'y a aucune inertie dans

[53] "Así, la adopción de la *política* de no detallar un compromiso es una política de 'autocubrimiento'. Adoptarla es, forzosamente, nunca hacerla explícita, 'ocultarla'" (Fingarette, 1969, p. 49).

[54] Rorty (1972), p. 392; de Sousa (1971), p. 312. Para la noción general implícita, *véase* C. Taylor (1971, 1976).

la conscience"),[55] convienen en que "dentro de ciertos límites, el decirlo lo hace verdad", para repetir una frase utilizada en el capítulo II, sec. 9. En este sentido, intencionalmente podemos adoptar nuevas creencias y actitudes, probándolas, por decirlo así, para ver si son compatibles con las condiciones limítrofes de nuestra vida. Desde luego, en algunos casos pueden resultar incompatibles; ninguna cantidad de esfuerzo podría cambiar a Garcin, de *Huis clos* ("A puertas cerradas", de Sartre) en un tipo distinto de persona. Garcin apela a la idea implícita en una frase citada en otra parte por Sartre,[56] "Je suis trop grand pour moi", manifiesto ejemplo de autoengaño. Sin embargo, en otros casos estas maniobras pueden resultar bien. Es posible superar el miedo repitiéndose a sí mismo "no tengo miedo", aunque inicialmente la afirmación sea falsa. El fracaso de la empresa no es condición suficiente para que hablemos de autoengaño en tales casos, pues esto implicaría que siempre podremos saber, de antemano, si resultará bien o no. Hay una "zona de indeterminación" dentro de la cual decir es hacer, y los límites de esta zona son indeterminados, es decir, incompletamente conocidos. Una vez que nos acostumbramos a este modo de considerar el mundo, sospecho que gran parte de nuestra experiencia puede embonar en este marco conceptual, de modo que muchos casos rudamente tildados de autoengaño deberán ser redescritos como impredecibles intentos fracasados de automodificación.

Algunos otros casos de autoengaño pueden ser interpretados por una variante de la teoría de Fingarette, de que el autoengaño es la negativa

[55] Sartre (1943), p. 101.
[56] *Ibid.*, p. 96.

a *declarar* nuestro compromiso en el mundo. La variedad de la declaración que yo tengo en mente es la reunión de toda prueba disponible, y el autoengaño surge cuando alguien deliberadamente se niega a recoger una prueba amenazadora (podemos observar, de paso, que la teoría de la racionalidad de que subraya el costo de la información está hecha a la medida como racionalización de esta actitud). El caso paradigmático es el dictador que dice a sus súbditos "No quiero conocer los detalles": aun cuando sepa que hay detalles desagradables que conocer, la falta de conocimiento específico le permite decirse a sí mismo y a los demás que no tiene conocimiento de que ocurran esas cosas. La habilidad de millones de alemanes para no enterarse de la exterminación de los judíos puede explicarse por este modelo: debieron de observar que sus conocidos judíos desaparecían, y debieron de saber que esto tenía alguna explicación horrible, pero mientras lograron permanecer ignorantes de los detalles, pudieron decirse a sí mismos que eran realmente ignorantes de lo que estaba ocurriendo. Este no es un caso puro o paradójico de autoengaño, porque no tenemos que imputar al autoengañador un conocimiento de los hechos que no quiere conocer, tan sólo un conocimiento de que hay tales hechos. La "Paradoja del prefacio" de poner en el prefacio de un libro que creemos que algunas de las afirmaciones hechas en el libro son falsas,[57] y la "paradoja de la calumnia", de decir que nos exponemos a las leyes contra la calumnia si expresamos nuestra verdadera opinión de una persona y que por tanto no lo haremos, también expresa esta distinción entre las afirmaciones específicas y las generales.

[57] *Cf.* Elster (1978a), p. 88.

En psicología cognoscitiva, esta variedad de autoengaño a menudo es llamada "exposición selectiva a la información",[58] creo que la teoría de Leon Festinger, que la disonancia cognoscitiva adquirió una falla fatal por no tomar en cuenta las implicaciones paradójicas de esta noción. Dado que la tesis general de la escuela de Festinger es que la gente trata de reducir la disonancia cognoscitiva, Festinger en su primera obra eligió como estrategia de particular importancia la exposición selectiva a la información, es decir, la tendencia a exponernos a fuentes informativas que producirán consonancia y evitarán que aumente la disonancia. La flagrante paradoja latente en esta formulación surgió en varios de los estudios experimentales utilizados para apoyar el argumento. Un estudio se refirió a la tendencia de los compradores de un producto, como un auto de cierta marca, a buscar y leer anuncios de esa marca aun *después* de la compra, y a evitar los anuncios de otras marcas. Aquí el problema es que no nos gusta hablar acerca de *información* en tales casos. La información, como el reconocimiento, debe proceder de una fuente independiente para que valga la pena. Buscar confirmación leyendo anuncios del producto que acabamos de comprar es como buscar el reconocimiento de un esclavo. En otro estudio, los resultados predichos por la teoría no se materializaron, y esto llevó a Festinger, en su obra posterior, a reconocer la paradoja: "Una vez que se ha dicho a la persona que la información que existe no apoya su decisión, se ha introducido ya evidencia adicional: en cierto sentido es imposible evitarla, pues sabe ya que exis-

[58] Lo siguiente se basa en *ibid.*, p. 88 *ss.*

te." [59] Si se hubiese hecho para empezar esta afirmación elemental, puede presumirse que habría sido diferente el diseño de toda la secuencia de experimentos. El descuido de la paradoja en los dos estudios citados equivale a la información implícita de que los sujetos se estaban engañando a sí mismos: que pudieron persuadirse de recabar información falsa de la cosa auténtica, y se contuvieron de elucidar los detalles. Ésta es, sin duda, una suposición interesante que hay que sondear, pero no es la hipótesis para la cual se idearon las pruebas.

Como breve conclusión, creo que las interpretaciones del autoengaño como automodificación y como negativa a recabar pruebas disponibles pueden explicar de modo no paradójico muchos casos importantes de mala fe. Repitiendo, no creo que puedan encargarse de todos los "casos difíciles", pero de todos modos contribuyen a una reducción del número de éstos. Sólo se logrará mayor progreso mediante nuevas reducciones.

[59] Festinger (1964), p. 82.

REFERENCIAS

Abelson, R. P. (1963), Computer simulation of hot cognition, S. Tomkins y S. Messik (comps.), *Computer simulation of personality*, pp. 277-298, Nueva York, Wiley.

Adserballe, H. (1977), *Frihedsberøvelse og trang i psykiatrien*. Copenhague, FADL's Forlag.

Ainslie, G. (1975), Specious reward. *Psychological Bulletin* 82, 463-496.

—— (1977), A behavioral understanding of the defense mechanism. Mimeografiado.

Alchian, A. (1950), Uncertainty, evolution and economic theory. *Journal of Political Economy* 58, 211-222.

Arrow, K. (1963), *Social choice and individual values*, 2ª ed. Nueva York, Wiley.

—— y Hurwicz, L. (1972), An optimality criterion for decision-marking under uncertainty, C. F. Carter y J. L. Ford (comps.), *Uncertainty and expectation in economics*, pp. 1-11, Clifton, N. J., Kelley.

Azzi, C., y Ehrenberg, R. (1975), Household allocation of time and church attendance, *Journal of Political Economy* 83, 27-56.

Bagehot, W. (1966), *The English constitution*, Ithaca, Nueva York, Cornell University Press.

Banton, M. (comp.) (1961), *Darwinism and the study of society*, Londres, Tavistock.

Barash, D. P. (1976), *Sociobiology and behavior*, Nueva York, Elsevier.

Barth, F. (1967), Economic spheres in Darfur, R. Firth (comp.), *Themes in economic anthropology*, pp. 149-174, Londres, Tavistock.

Baumol, W. (1952), *Welfare economic and the theory of the state*, Londres, Bell.

Becker, G. (1976), *The economic approach to human behavior*, Chicago, Chicago University Press.

Bellman, R. (1961), *Adaptive control processes*, Princeton, N. J., Princeton University Press.

Bhaduri, A. (1973), A study in agricultural backwardness under semi-feudalism, *Economic Journal* 83, 120-137.

Bohannan, P. (1955), Some principles of exchange and investment among the Tiv, *American Anthropologist*, nueva serie, 57, 60-70.

Borch, K. (1968), *The economics of uncertainty*, Princeton, N. J., Princeton University Press.

Boudon, R. (1973a), *L'inégalité des chances*, París, Armand Colin.

—— (1973b), *Mathematical structures of social mobility*, Amsterdam, Elsevier.

—— (1974), *Education, opportunity and social inequality*, Nueva York, Wiley.

—— (1977), *Effets pervers et ordre social*, París, Press Universitaires de France.

Bourdieu, P., y Passeron, J.-C. (1967), *Les héritiers*, París, Editions de Minuit, 1970. *La reproduction*, París, Editions de Minuit.

—— y Chamoredon, J.-C. (1968), *Le métier de sociologue*, La Haya, Mouton.

Bowles, S., y Gintis, H. (1976), *Schooling in capitalist America*, Londres, Routledge and Kegan Paul.

Brams, S. (1976), *Paradoxes in politics*, Nueva York, Free Press.

Brooke, M. (1972), Problems in the decision-making process. M. Brooke y H. Remmers (comps.) *The multinational company in Europe*, pp. 93-109, Londres, Longman.

Cargile, J. (1966), Pascal's wager, *Philosophy* 35, 250-257.

Clements, A. L. (comp.) (1966), *John Donne's poetry*, Nueva York, Norton.

Cobb, W. (1973), Theft and the two hypotheses. S. Rottenberg (comp.), *The economics of crime and punishment*, pp. 19-30, Washington, D. C., American Enterprise Institute for Public Policy Research.

Coddington, A. (1968), *Theories of the bargaining process*, Londres Allen and Unwin.

Cody, M. L. (1974), Optimization in ecology, *Science* 183, 1156-1164.

Cohen, S. N. (1977), Recombinant DNA: fact and fiction, Science 195, 654-657.
Coser, L. (1971), Social conflict and the theory of social change, C. G. Smith (comp.), *Conflict resolution: contributions of the behavioral sciences*, páginas 58-65, Notre Dame, Ind., University of Notre Dame Press.
Cotgrove, S. y Box, S. (1970), *Science, industry and society*, Londres, Allen and Unwin.
Curio, E. (1973), Towards a methodology of teleonomy, *Experientia* 29, 1045-1058.
—— (1976), *The ethology of predation*, Berlín, Springer.
Cyert, R. M., y De Groot, M. H. (1975), Adaptive utility, R. H. Day, y T. Groves (comps.), *Adaptive economic models*, pp. 223-246, Nueva York, Academic Press.
Darlington, C. D. (1969), *The evolution of man and society*, Nueva York, Simon and Schuster.
David, P. (1975), *Technical choice, innovation and economic growth*, Cambridge, Cambridge University Press.
Davidson, D. (1969), How is weakness of the will possible? J. Feinberg (comp.), *Moral concepts*, pp. 93-113. Oxford, Oxford University Press. [Hay versión en español, FCE.]
—— (1970), Mental events. L. Foster, y J. W. Swanson (comps.), *Experience and theory*, pp. 79-101. Amherst, Mass., University of Massachusetts Press.
—— (1973), The material mind. P. Suppes *et al.* (comps.), *Logic, methodology and philosophy of science IV*, pp. 709-724. Amsterdam, Holanda.
—— (1976), Psychology as philosophy. J. Glover (comp.), *The philosophy of mind*, pp. 101-110, Oxford, Oxford University Press.
Dawkins, R. (1976), *The selfish gene*, Nueva York, Oxford University Press.
Dennett, D. (1971), Intentional systems, *Journal of Philosophy* 68, 87-106.
—— (1976), Conditions of personhood. A. Rorty (comp.), *The identifies of persons*, pp. 175-196, Berkeley, University of California Press.

Descartes, R. (1897-1910), *Oeuvres complètes*, C. Adam y P. Tannery (comps.), 11 vols. París, Vrin.
Dickinson, Emily (1970), *Complete poems*, Thomas H. Johnson (comp.), Faber and Faber.
Dorfman, R., Samuelson, P. y Solow, R. (1958), *Linear programming and economic analysis*, Nueva York, McGraw-Hill.
Dworkin, G. (1972), Paternalism, *The Monist* 56, 64-84.
Edmunds, M. (1974), *Defence in animals*, Harlow, Essex, Longman.
Ellsberg, D. (1971), The quagmire myth and the stalemate machine, *Public Policy* 19, 217-274.
Elster, J. (1975a), *Leibniz et la formation de l'esprit capitaliste*, París, Aubier.
—— (1075b), Review of Francis Sejersted, *Ideal, teori og virkelighet*, *Economic History Review* 2ª ser. 28, 160.
—— (1976a), A note on hysteresis in the social sciences, *Synthese* 33, 371-391.
—— (1976b), Some conceptual problems in political theory. B. Barry (comp.), *Power and political theory*, pp. 245-270. Londres, Wiley.
—— (1978a), *Logic and society*, Londres, Wiley.
—— (1978b), Exploring exploitation, *Journal of Peace Research* 15, 3-17.
—— (1978c), The labor theory of value, *Marxist Perspectives* 1, 70-101.
Evans-Pritchard, E. (1940), *The Nuer*, Oxford, Oxford University Press.
Fararo, T. S. (1973), *Mathematical sociology*, Nueva York, Wiley.
Farber, L. (1976), *Lying, despair, jealousy, envy, sex, suicide, drugs and the good life*, Nueva York, Basic Books.
Feller, W. (1968), *An introduction to probability theory and its applications*, vol I, 3ª ed. Nueva York, Wiley.
Fellner, W. (1961), Two propositions in the theory of induced innovations, *Economic Journal* 71, 305-308.
—— (1965), *Probability and profits*, Homewood, Ill., Irwin.

Festinger, L. (1964), Conflict, decision and dissonance, Stanford, Calif., Stanford University Press.
Fingarette, H. (1969), *Self-deception*, Londres, Routledge and Kegan Paul.
Finley, M. I. (1973), *Democracy: ancient and modern*, Londres, Chatto and Windus.
Fishburn, P. (1973), On the foundations of game theory: the case of non-Archimedean Utilities, *International Journal of Game Theory* 2, 65-71.
Foreyt, J. P. (comp.) (1977), *Behavioral treatment of obesity*, Oxford, Pergamon Press.
Frankfurt, H. M. (1971), Freedom of will and the concept of a person, *Journal of Philosophy* 68, 5-20.
Frazzetta, T. H. (1975), *Complex adaptations in evolving populations*, Sunderland, Mass., Sinauer.
Freud, S. (1969), *Gesammelte Werke*, Frankfurt a. M., Fischer.
Friedman, M. (1953), The methodology of positive economics. M. Friedman, *Essays in positive economics*, pp. 3-43, Chicago, University of Chicago Press.
Gardiner, P. (1969-1970), Error, faith and self-deception, *Proceedings of the Aristotelian Society*, nueva serie, 70, 197-220.
Genovese, E. (1965), *The political economy of slavery*, Nueva York, Pantheon.
—— (1974), *Roll, Jordan, Roll*, Nueva York, Phanteon.
Georgescu-Roegen, N. (1950), The theory of choice and the constancy of economic laws, *Quarterly Journal of Economics* 64, 125-138.
—— (1954), Choice, expectations and measurability, *Quarterly Journal of Economics* 68, 503-534.
Gerschenkron, A. (1970), *Europe in the Russian mirror*, Cambridge, Cambridge University Press.
Gibbard, A. (1974), A Pareto-consistent libertarian claim, *Journal of Economic Theory* 7, 388-410.
Glantz, S. A., y Albers, N. V. (1974), Department of Defense R & D in the university, *Science* 186, 706-711.
Goldman, A. (1972), Toward a theory of social power, *Philosophical Studies* 23, 221-268.

Gorman, W. M. (1967), Tastes, habits and choices, *International Review* 8, 218-222.
Gould, S. J. (1976), D'Arcy Thompson and the science of form. M. Greene y E. Mendelsohn (comps.), *Topics in the philosophy of biology*, pp. 66-97, Dordrecht, Reidel.
Griffin, D. R. (1976), *The question of animal awareness*, Nueva York, Rockefeller University Press.
Grindheim, S. (1975), Hvorfor fikk aristokratiet regjere når borgerskapet hersket? Tesis de maestría en historia, Universidad de Oslo, mimeografiado.
Gullvåg, I. (1977), The logic of assertion, University of Trondheim, mimeografiado.
Haavelmo, T. (1944), The probability approach in econometrics, *Econometrica* (suplemento) 12, 1-118.
—— (1970), Some observations on welfare and economic growth. W. A. Eltis, M. Scott y N. Wolfe (comps.), *Induction, growth and trade: essays in honour of Sir Roy Harrod*, pp. 65-75, Oxford, Oxford University Press.
Hacking, I. (1975), *The emergence of probability*, Cambridge, Cambridge University Press.
Hagenbüchle, R. (1974), Precision and indeterminacy in the poetry of Emily Dickinson, *Emerson Society Quarterly* 20, 33-56.
Hájek, O. (1975), *Pursuit games*, Nueva York, Academic Press.
Hamilton, W. (1964), The genetical theory of social behaviour, *Journal of Theoretical Biology* 7, 1-52.
—— (1967), Extraordinary sex ratios, *Science* 156, 477-488.
Hammond, P. (1976), Changing tastes and coherent dynamic choice, *Review of Economic Studies* 43, 159-173.
Harsanyi, J. (1955), Cardinal welfare, individualistic ethics and interpersonal comparisons of utility, *Journal of Political Economy* 63, 309-321
—— (1976), Can the maximin principle serve as a basis for morality? *American Political Science Review* 69, 594-606.
—— (1977), *Rational behavior and bargaining equilib-*

rium in games and social situations, Cambridge, Cambridge University Press.

Heal, G. (1973), *The theory of economic planning*, Amsterdam, Holanda.

Hegel, G. W. F. (1807), *Die Phänomenologie des Geistes*, Hamburgo, Felix Meiner, 1953.

Henry, C. (1974), Investment decisions under uncertainty: the "irreversibility effect", *American Economic Review* 64, 1106-1012.

Hicks, J. (1932), *The theory of wages*, Londres, Macmillan.

Himes, J. H. (1971), The functions of racial conflict. C. G. Smith (comp.), *Conflict resolution*, pp. 170-179, Notre Dame, Ind., University of Notre Dame Press.

Hintikka, J. (1967), *Cogito ergo sum:* inference or performance? W. Doney (comp.), *Descartes*, pp. 108-139, Londres, Macmillan.

Hirsch, F. (1976), *Social limits to growth*, Cambridge, Mass., Harvard University Press.

Hirschman, A. O. (1971), *Exit, voice and loyalty*, Cambridge, Mass., Harvard University Press. [Hay versión en español del FCE.]

—— (1977), *The passions and the interests*, Princeton, N. J. Princeton University Press. [Hay versión en español del FCE.]

Hobsbawm, E. (1959), *Primitive rebels*, Manchester, Manchester University Press.

Howard, N. (1971), *Paradoxes of rationality*, Cambridge, Mass., MIT Press.

Husserl, E. (1966), *Zur Phänomenologie des inneren Zeitbewusstseins* (= Husserliana, vol. x), La Haya, Martinus Nijhoff.

Isaacs, R. (1965), *Differential games*, Nueva York, Wiley.

Jacob, F. (1970), *La logique du vivant*, París, Gallimard.

James, W. (1896), The will to believe. *The will to believe and other essays in popular philosophy*, Nueva York, Dover, 1897.

Janis, I. (1972), *Victims of group-think*, Boston, Houghton Mifflin.

——, y Mann, L. (1977), *Decision making*, Nueva York, Free Press.

Jarvie, I. C. (1968), Limits to functionalism and alternatives to it in anthropology. R. A. Manners y D. Kaplan (comps.), *Theory in anthropology*, pp. 196-203, Londres, Routledge and Kegan Paul.

Jeffrey, D. B. (1975), Self-control: methodological issues and research trends. M. J. Mahoney y C. E. Thoresen (comps.) (1975), pp. 166-199.

—— (1977), Introduction to part VI of Foreyt (comp.) (1977).

Jewell, D. (1977), *Duke: a portrait of Duke Ellington*, Nueva York, Norton.

Johansen, L. (1973), On the optimal use of forecasts in public policy, *Journal of Public Economics* 1, 1-24.

—— (1977), *Lectures on macro-economic planning*, vol. I, Amsterdam, Holanda.

Kalecki, M. (1971), Political aspects of full employment. M. Kalecki, *Selected essays on the dynamics of the capitalist economy*, pp. 138-145, Cambridge, Cambridge University Press.

Keeney, R. L. y Raiffa, H. (1976), *Decisions with multiple objectives*, Nueva York, Wiley.

Kennedy, C. (1964), Induced innovation and the theory of distribution, *Economic Journal* 74, 541-547.

Kydland, F. y Prescott, E. (1977), Rules rather than discretion: the inconsistency of optimal plans, *Journal of Political Economy* 85, 473-492.

Lancaster, K. (1973), The dynamic inefficiency of capitalism, *Journal of Political Economy* 81, 1092-1109.

Latsis, S. (1976), A research programme in economics. S. Latsis (comp.), *Methods and appraisal in economics*, pp. 1-41, Cambridge, Cambridge University Press.

Laudan, L. (1977), *Progress and its problems*, Berkeley, University of California Press.

LeClair, E. y Schneider, H. (comps.) (1968), *Economic*

anthropology, Nueva York, Holt, Rinehart and Winston.

Leibniz, G. W. (1875-1890), *Die philosophische Schriften*. C. J. Gerhardt. (comp.), 6 vols., Berlín, Weidmannsche Buchhandlung.

Leigh, E. G. (1971), *Adaptation and diversity*, San Francisco, Freeman and Cooper.

Lindbeck, A. (1976), Stabilization policy in open economies with endogenous politicians, *American Economic Review: Papers and Proceedings* 66, 1-19.

Luce, R. D. y Raiffa, H. (1957), *Games and decisions*, Nueva York, Wiley.

——, y Suppes, P. (1965), Preference, utility and subjective probability. R. D. Luce, R. R. Bush y E. Galanter (comps.), *Handbook of mathematical psychology*, vol. III, pp. 249-410, Nueva York, Wiley.

McFarland, A. S. (1969), *Power and leadership in pluralist systems*, Stanford, Calif., Stanford University Press.

Mahoney, M. J. (1972), Research issues in self-management, *Behavior Therapy* 3, 45-63.

——, y Thoresen, C. E. (comps.) (1975), *Self-control: power to the person*, Monterey, Calif., Brooks/Cole.

Mangset, D. (1974), Industri, stat og karteller. Tesis de maestría en historia, Universidad de Oslo. Mimeografiado.

March, J. (1977), Bounded rationality, ambiguity and the engineering of choice. Stanford University, Calif. Mimeografiado.

Marsily G. de *et al.* (1977), Nuclear waste disposal, *Science* 197, 519-527.

Marx, K. (1845-1846), *Die deutsche Ideologie*. Marx-Engels Werke, vol. III. Berlín, Dietz, 1962.

—— (1852), *Louis Napoleon's 18 Brumaire*.

—— (1857-1858), *Grundrisse der Kritik der politischen Ökonomie*, Berlín: Dietz, 1953.

—— (1867), *Capital*, vol. I, tr. S. Moore y E. Aveling, Nueva York, International Publishers, 1967.

—— (1971), *The unknown Karl Marx*, (comp.), R. Payne, Nueva York, New York University Press.

Marx, K. y Engels, F. (1848), *El manifiesto comunista.*

—— (1971), *Articles on Britain,* Moscú, Progress Publishers.

Masters, R. D. (1975), Politics as a biological phenomenon. *Social Science Information* 14, 7-63.

—— (1976), Functional approaches to analogical comparisons between species. M. von Cranach (comp.), *Methods of inference from animal to human behaviour,* pp. 73-102. La Haya, Mouton.

Maynard Smith, J. (1973), The logic of animal conflict, *Nature, London* 246, 15-18.

—— (1974), The theory of games and the evolution of animal conflict, *Journal of Theoretical Biology* 47, 209-221.

Mayr, E. (1970), *Populations, species and evolution,* Cambridge, Mass., Harvard University Press.

Meisner, M. (1967), *Li Tao-chao and the origins of Chinese communism,* Cambridge, Mass., Harvard University Press.

Merton, R. K. (1957), *Social theory and social structure,* Glencoe, Ill., Free Press.

—— (1973), *The sociology of science,* Chicago, University of Chicago Press.

Millay, Edna St. Vincent (1975), *Collected poems,* (comp.), N. Millay, Nueva York, Harper and Row.

Miller, D. (1976), *Social justice,* Oxford, Oxford University Press.

Monod, J. (1970), *Le hasard et la nécessité,* París, Le Seuil.

Montmort, R. de (1713), *Essai d'analyse sur les jeux de hazard.* 2ª ed. París.

Morgenstern, O. (1964), Pareto-optimum and economic organization. *Systeme und Methoden in den Wirtschafts- und Sozialwissenschaften: Festschrift für Erwin von Beckerath,* pp. 573-586. Tubinga, Mohr.

—— (1972), Thirteen critical points in contemporary economic theory, *Journal of Economic Literature* 10, 1163-1189.

—— (1976), Some reflections on utility. A. Schotter (comp.), *Selected economic writings of Oskar Mor-*

genstern, pp. 65-70. Nueva York, New York University Press.
Morishima, M. y Seton, F. (1961), Aggregation in Leontief matrices and the labour theory of value, *Econometrica* 29, 203-220.
Mortimore, G. W. (comp.) (1971), *Weakness of will*, Londres, Macmillan.
Nagel, T. (1970), *The possibility of altruism*, Oxford, Oxford University Press.
Needham, J. (1956), *Science and civilisation in China*, vol. II, Cambridge, Cambridge University Press.
Nelson, R. y Winter, S. G. (1974), Neoclassical vs. evolutionary theories of economic growth. *Economic Journal* 84, 886-905.
—— (1976), Technical change in an evolutionary model. *Quarterly Journal of Economics* 90, 90-118.
Newton-Smith, W. (1973), A conceptual investigation of love. A. Montefiore (comp.), *Philosophy and personal relations*, pp. 113-135, Londres, Routledge and Kegan Paul.
Nivison, D. (1976), Motivation and moral action in Mencius. Stanford University, Calif. Mimeografiado.
Nordhaus, W. (1975), The political business cycle. *Review of Economic Studies* 42, 169-190.
North, D. (1971), Institutional change and economic growth. *Journal of Economic History* 31, 118-125.
—— y Thomas, R. (1973), *The rise of the western world*, Cambridge, Cambridge University Press.
Nozick, R. (1969a), Coercion. S. Morgenbesser *et al.* (comps.), *Philosophy, science and method: essays in honor of Ernest Nagel*, pp. 440-472, Nueva York, St. Martin's Press.
—— (1969b), Newcomb's Problem and two principles of choice. N. Rescher (comp.), *Essays in honor of Carl G. Hempel*, pp. 114-146, Dordrecht, Reidel.
—— (1974), *Anarchy state and utopia*, Oxford, Blackwell.
Olson, M. (1965), *The logic of collective action*, Cambridge, Mass., Harvard University Press.
Owen, G. (1968), *Game theory*, Filadelfia, Saunders.

Padulo, L., y Arbib, M. (1974), *System theory*, Filadelfia, Saunders.

Parfit, D. (1973), Later selves and moral principles. A. Montefiore (comp.), *Philosophy and personal relations*, pp. 137-169, Londres, Routledge and Kegan Paul.

Pears, D. (1974), Freud, Sartre and self-deception. R. Wollheim (comp.), *Freud*, pp. 97-112, Nueva York, Anchor Books.

Peston, M. H. (1967), Changing utility functions. M. Shubik (comp.), *Essays in mathematical economics in honor of Oskar Morgenstern*, pp. 233-236, Princeton, N. J., Princeton University Press.

Peyre, H. (1963), *Literature and sincerity*, New Haven, Conn., Yale University Press.

—— (1967), *The failures of criticism*, Ithaca, Nueva York, Cornell University Press.

Phelps, E. S., y Pollak, R. A. (1968), On second-best national saving and game-theoretic equilibrium growth, *Review of Economic Studies* 35, 185-199.

Pollak, R. A. (1968), Consistent planning, *Review of Economic Studies* 35, 201-208.

—— (1970), Habit formation and dynamic demand functions, *Journal of Political Economy* 78, 745-763.

—— (1976), Habit formation and long-run preferences, *Journal of Economic Theory* 13, 272-297.

Popper, K. (1957), *The poverty of historicism*, Londres, Routledge and Kegan Paul.

—— (1972), *Objective knowledge*, Oxford, Oxford University Press.

Quirk, J., y Saposnik, R. (1968), *Introduction to general equilibrium theory and welfare economics*, Nueva York, McGraw-Hill.

Raiffa, H. (1968), *Decision analysis*, Reading, Mass., Addison-Wesley.

Rapoport, A. (1966), *Two-person game theory*, Ann Arbor, University of Michigan Press.

——, y Chammah, A. (1965), *Prisoner's Dilemma*, Ann Arbor, University of Michigan Press.

——, Guyer, M. J. y Gordon, D. C. (1976), The 2×2 *game*, Ann Arbor, University of Michigan Press.

Rapport, D. J., y Turner, J. E. (1977), Economic models in ecology, *Science* 195, 367-373.
Rawls, J. (1971), *A theory of justice*, Cambridge, Mass., Harvard University Press.
Reid, C. (1970), *Hilbert: the life and story of David Hilbert*, Berlín, Springer.
Reisner, R. (1974), *Bird: the legend of Charlie Parker*, Londres, Quartet Books.
Richerson, P. J., y Boyd, R. A dual inheritance model of the human evolutionary process I, *Journal of Human and Social Biology*.
Rigby, P. W. J., Burleigh, B. D. y Hartley, B. S. (1974), Gene duplication in experimental enzyme evolution, *Nature, London* 251, 200-204.
Riker, W., y Ordeshook, P. C. (1973), *An introduction to positive political theory*, Englewood Cliffs, N. J., Prentice-Hall.
Robbins, S. M., y Stobaugh, R. B. (1974), *Money in the multinational enterprise*, Harlow, Essex, Longman.
Robinson, J. (1956), *The accumulation of capital*, Londres, Macmillan.
Rorty, A. (1972), Belief and self-deception, *Inquiry* 15, 387-410.
Rosen, R. (1972), Morphogenesis. R. Rosen (comp.), *Foundations of mathematical biology*, vol. II, páginas 1-77, Nueva York, Academic Press.
Rosenberg, N. (1976), *Perspectives on technology*, Cambridge, Cambridge University Press.
Roumasset, J. A. (1976), *Rice and risk: decision-making among low-income farmers*, Amsterdam, North Holland.
Ruse, M. (1974), Cultural evolution, *Theory and Decision* 5, 413-440.
Russell, R. (1973), *Bird lives: the high life and hard times of Charlie Parker*, Nueva York, Charterhouse.
Ryle, G. (1949), *The concept of mind*, Londres, Hutchinson.
Salter, W. E. G. (1960), *Productivity and technical change*, Cambridge, Cambridge University Press.

Salthe, S. (1972), *Evolutionary biology*, Nueva York, Holt, Rinehart and Winston.
Sanders, W. (1971), *John Donne's poetry*, Cambridge, Cambridge University Press.
Sartre, J.-P. (1943), *L'être et le néant*, París, Gallimard.
Schafer, R. (1976), *A new language for psychoanalysis*, New Haven, Conn., Yale University Press.
Schelling, T. S. (1963), *The strategy of conflict*, Cambridge, Mass., Harvard University Press.
Schlanger, J. (1971), *Les métaphores de l'organisme*, París. Vrin.
Schoffeniels, E. (1973), *L'anti-hasard*, París, Gauthier-Villars.
Schumpeter, J. (1934), *The theory of economic development*, Cambridge, Mass., Harvard University Press.
—— (1954), *A history of economic analysis*, Londres, Allen and Unwin.
Sejersted, F. (1973), *Ideal, teori og virkelighet*, Oslo, Cappelen.
—— (1978), Rettsstaten og den selvdestruerende makt. R. Slagstad (comp.), *Om staten*, pp. 46-84, Oslo, Pax.
Sen, A. K. (1967), Isolation, assurance and the social rate of discount, *Quarterly Journal of Economics* 80, 112-124.
—— (1970), *Collective choice and social welfare*, San Francisco, Holden Day.
—— (1973), *On economic inequality*, Oxford, Oxford University Press.
—— (1974), Choice, orderings and morality. S. Körner (comp.), *Practical reason*, pp. 54-67, Oxford, Blackwell.
—— (1975), *Employment, technology and development*, Oxford, Oxford University Press.
Shapley, L., y Shubik, M. (1967), Ownership and the production function *Quarterly Journal of Economics* 80, 88-111.
Shefrin, H., y Thaler, R. (1977), An economic theory

of self-control, University of Rochester, Nueva York. Mimeografiado.

Shubik, M. (1970), Game theory, behavior and the paradox of the Prisoner's Dilemma, *Journal of Conflict Resolution* 14, 181-202.

Simon, H. (1954), A behavioral theory of rational choice, *Quarterly Journal of Economics* 69, 99-118.

—— (1976), From substantive to procedural rationality. S. Latsis (comp.), *Method and appraisal in economics*, pp. 129-148, Cambridge, Cambridge University Press.

Simon, M. (1971), *The matter of life*, New Haven, Conn., Yale University Press.

Snell, J. L. (1965), Stochastic processes. R. D. Luce, R. Bush y E. Galanter (comps.), *Handbook of mathematical psychology*, vol. III, pp. 411-486, Nueva York, Wiley.

Sousa, R. de (1971), Review of Fingarette (1969), *Inquiry* 14, 308-321.

Stanford Research Institute (1976), The economic and social costs of coal and nuclear electric generation. Report prepared for the National Science Foundation, Stanford, Calif.

Stark, W. (1962), *The fundamental forms of social thought*, Londres, Routledge and Kegan Paul.

Stebbins, G. L., y Lewontin, R. C. (1972), Comparative evolution at the levels of molecules, organisms and populations. L. M. Le Cam, J. Neyman y E. L. Scott (comps.), *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*, vol. V, pp. 23-42, Berkeley, University of California Press.

Stendhal (1952), *Romans et nouvelles*. H. Martineau (comp.). 2 vols., París, Gallimard (Editions de la Pléiade).

Stigler, G., y Becker, G. (1977), De gustibus non est disputandum, *American Economic Review* 67, 76-90.

Stinchcombe, A. (1968), *Constructing social theories*, Nueva York, Harcourt, Brace and World.

Strotz, R. H. (1955-1956), Myopia and inconsistency

in dynamic utility maximisation, *Review of Economic Studies* 23, 165-180.
Stuart, R. B. (1971), A three-dimensional program for the treatment of obesity, *Behaviour Research and Theory* 9, 177-186.
Suttmeier, R. (1974), *Research and revolution*, Lexington, Mass., Lexington Books.
Taylor, A. J. (1960), Progress and poverty in Britain 1750-1850, *History* 45, 16-31.
Taylor, C. (1971), Interpretation and the sciences of man, *Review of Metaphysics* 25, 3-51.
—— (1976), Responsibility for self. A. Rorty (comp.), *The identities of persons*, pp. 281-300, Berkeley, University of California Press.
Taylor, G. (1975-1976), Love, *Proceedings of the Aristotelian Society*, nueva serie 76, 147-164.
Taylor, M. (1973), The problem of salience in the theory of collective decision-making. H. R. Alker, K. W. Deutsch y A. H. Stoetzel (comps.), *Mathematical approaches to politics*, pp. 231-256, Amsterdam, Elsevier.
—— (1976a), *Anarchy and cooperation*, Londres, Wiley.
—— (1976b), Altruism and cooperation: their evolution by natural selection and their role in primitive societies, University of Essex. Mimeografiado.
Tisdell, C. (1971), Economic policy, forecasting and flexibility, *Weltwirtschaftliches Archiv* 106, 34-51.
Trivers, R. (1971), The evolution of reciprocal altruism, *Quarterly Review of Biology* 46, 35-57.
—— (1972), Parental investment and sexual selection. B. Campbell (comp.), *Sexual selection and the descent of man*, pp. 136-179, Chicago, Aldine.
Tullock, G. (1974), Does punishment deter crime? *The Public Interest* 36, 103-111.
Tversky, A., y Kahneman, D. (1974), Judgement under uncertainty, *Science* 185, 1124-1130.
Ulam, S. (1976), *Adventures of a mathematician*, Nueva York, Scribner.
Valéry, P. (1957), Stendhal. *Oeuvres*, vol. I, pp. 553-582, París, Gallimard (Editions de la Pléiade).

Watzlawick, P. (1977), *How real is real?* Nueva York, Vintage Books.
——, Beavin, J. H. y Jackson, D. D. (1967), *Pragmatics of human communication,* Nueva York, Norton.
——, Weakland, C. E. y Fisch, R. (1974), *Change,* Nueva York, Norton.
Weiszäcker, C. C. von (1971), Notes on endogenous change of tastes, *Journal of Economic Theory* 3, 345-372.
—— (1973), Notes on endogenous growth of productivity. J. A. Mirrlees y N. H. Stern (comps.), *Models of economic growth,* pp. 101-114, Londres, Macmillan.
Wharton, C. (1971), Risk, uncertainty and the subsistence farmer. G. Dalton (comp.), *Economic development and social change,* pp. 566-574, Nueva York, Natural History Press.
Williams, B. A. O. (1973), Deciding to believe. *Problems of the self,* pp. 136-151, Cambridge, Cambridge University Press.
—— (1976), Persons, character and morality. A. Rorty (comp.), *The identities of persons,* pp. 197-216, Berkeley, University of California Press.
Williams, G. C. (1966), *Adaptation and natural selection,* Princeton, N. J., Princeton University Press.
Williams, J. D. (1954), *The compleat strategyst,* Nueva York, McGraw-Hill.
Wilson, E. O. (1975), *Sociobiology,* Cambridge, Mass., Harvard University Press.
Winter, S. G. (1964), Economic "natural selection" and the theory of the firm, *Yale Economic Essays* 4, 225-272.
—— (1971), Satisficing, selection and the innovating remnant, *Quarterly Journal of Economics* 85, 237-261.
—— (1975), Optimization and evolution. R. H. Day y T. Groves (comp.), *Adaptive economic models,* pp. 73-118, Nueva York, Academic Press.
Wittgenstein, L. (1953), *Philosophische Untersuchungen,* Oxford, Blackwell.

Wood, M. (1971), *Stendhal*, Ithaca, Nueva York, Cornell University Press.

Wright, S. (1970), Random drift and the shifting balance theory of evolution. K. Kojima (comp.), *Mathematical topics in population genetics* (= *Biomathematics*, vol. I), pp. 1-31, Berlín, Springer.

ÍNDICES

ÍNDICE ANALÍTICO

abdicación: 95, 108, 154, 164, 167 *ss*
Ainslie, G.: 66, 74, 77, 131, 146 *ss*, 162, 177 *n*, 181 *n*, 183, 185 *n*
akrasia, *véase* voluntad
aleatoriedad: 29, 155 *n*, 202, 233, 256
altruismo: 50 *ss*, 112 *ss*, 235; condicional: 44; incondicional: 44; recíproco: 241
amor: 247, 249, 261, 273 *ss*
apuestas:
  privadas indirectas: 78, 175, 176
  públicas indirectas: 174, 177
Aristóteles: 92, 99, 143, 178, 180, 292
Asamblea Constituyente: 159, 174
Azar, *véase* aleatoriedad

Becker, G.: 50, 51, 54, 113, 194 *n*, 212 *n*, 228, 239, 240, 242, 245 *n*, 249 *n*
Blake, W.: 265
Boudon, R.: 39, 47, 230

Cambio técnico: 18 *n*, 27 *n*, 29, 206 *ss*, 257
Capitalismo: 156, 159, 163 *ss*, 268 *n*
caridad: 197, 256 *ss*
Catch 22: 184, 252, 261, 263
coerción: 140, 142
contradicción: 159, 198, 249 *ss*, 261, 266, 270, 288, 291
Creencia: 84 *ss*, 175, 176, 218, 260
Curio, E.: 14 *n*, 32, 33 *n*, 46 *n*

David, P.: 18 *n*, 27 *n*, 28 *n*, 206 *n*
Davidson, D.: 11, 88 *n*, 92 *n*, 254, 255, 289 *n*
Dawkins, R.: 49, 240
decisión de creer: 84 *ss* 88, 250, 251, 252, 292
  decidir: 76
  olvidar: 88 *n*, 251, 269
Dennett, D.: 36
Descartes, R.: 81, 95 *ss*
Dickinson, E.: 88 *n*, 142 *n*, 251, 265, 279
Dilema del preso: 40, 42, 164, 203, 237 *ss*, 280
  doble atadura: 89 *n*, 261, 266, 269
Donne, J.: 119 *n*, 262, 264, 265, 271, 275, 278, 281, 293
elecciones: 107, 149, 154, 155
equilibrio: 39, 199, 238
estructuralismo: 192
estrategias:

indirecta: 23 *ss*, 78, 79, 148
mixta: 46, 201, 234, 248
explicación de la mano invisible: 57
explicaciones de filtro: 57, 228

Farber, L.: 251, 253
Fellner, W.: 205 *n*, 219
Festinger, L.: 266, 296
Fingarrette, H.: 287 *ss*
Finley, M. I.: 151 *ss*
Fishburn, P.: 209 *n*
Frankfurt, H.: 187
Freud, S.: 95, 185, 265, 266, 272, 286
Friedman, M.: 105, 223
funcionalismo: 54 *ss*, 81, 103, 152, 173

Gallinas: 205, 207
Genovese, E.: 194, 270
grafe paranomon: 151
Griffin, D.: 36

Hamilton, W.: 52, 54
Hammond, P.: 116, 118 *n*
Harsanyi, J.: 194 *n*, 200 *ss*, 215 *n*, 234 248
Hegel, G.: 150, 171, 253, 261, 267 *ss*, 276 *n*
Hirschman, A.: 97, 108 *n*
histéresis: 245
Husserl, E.: 273 *n*, 282 *n*

Idea reguladora: 13, 14, 62
incertidumbre: 114, 221
información: 42, 44, 101 *ss*, 175, 223, 226, 258, 277, 296
intencionalidad: 11, 35, 37, 152, 233, 254, 272, 289
inversión: 22, 25, 75, 80
irracionalidad: 117 *ss*, 244 *ss*, 248 *ss*, 256, 260

Janis, I.: 266, 287, 290
Jarvie, I. C.: 61
Juego del seguro: 42 *ss*, 242
juegos: sin estrategias dominantes: 38, 42, 50, 53, 242, 275, 276
sin puntos de equilibrio: 200
sin soluciones: 197 *ss*, 276

Kalecki, M.: 65
Kant, I: 13 *n*
Kydland, F.: 155 *ss*, 165

Lancaster, K.: 156, 157 *n*
Leibniz, G.: 24, 39, 101, 102, 103, 158 *n*, 207, 247, 281
Lindbeck, A.: 108, 154

Mann, L.: 266, 287 *n*, 290
Marx, G.: 253, 261, 275
Marx, K.: 39, 164 *ss*, 211, 268 *n*
marxista: 56, 63, 65, 211
maximizar: 101, 223, 224 *ss*
global: 12, 13, 14, 23 *ss*, 148
local: 13, 16 *ss*, 55
Maynard Smith, J. 49, 54

Merton, R.: 15, 54 *ss*
miopía: 24, 28, 147, 148
Morgenstern, O.: 189
mutaciones: 17 *ss*, 27

Nagel, T.: 73, 117 *n*
Neumann, J. von: 47, 197 *n*, 291
Nordhaus, W.: 106, 153
North, D.: 102, 194, 258
Nozick, R.: 57, 142, 245 *n*. 284

odio: 261, 266
ostracismo: 151

paradojas: 248 *ss*
  paradoja del examen: 281 *n*
  paradoja de evitar información: 176 *n*, 296
  paradoja del Prefacio: 295
  paradoja de la calumnia: 295
Parfit, D.: 117 *n*, 184, 252
Pascal, B.: 84 *ss*, 99, 114
paternalismo. 83 *n*, 144
persuasión: 140, 142, 144
Peyre, H.: 183 *n*, 253 *n*, 275, 290
Phelps, E. S.: 124, 128 *n*
Pollak, R.: 124, 126, 128, 133 *n*, 134 *n*, 144
Popper, K.: 26 *n*
preferencias: no arquimédicas: 212, 213, 243
  cambiantes: 116, 134 *ss*, 232, 244 *ss*, 258
  intransitivas: 247
  lexicográficas: 208 *ss*, 229, 236 *n*
  temporales: 115 *ss*
Prescott, E.: 155 *ss*, 165
probabilidades subjetivas: 214 *ss*
problema de Newcomb: 88 *n*, 284
Punto fijo: 85, 162, 179 *n*
racionalidad: colectiva, 39, 41; primera mejor: 244; imperfecta: 66, 188, 286; individual: 39; instantánea: 96, 99, 109; limitada: 101, 224; paramétrica: 38, 197; perfecta: 11; presunción de: 149, 254; segunda mejor: 188, 244; estratégica: 38, 199 *ss*; temporal: 115, total: 101
  racionalización de los deseos: 289, 290
Rapoport, A.: 40, 204, 236
Rawls, J.: 117, 203
reduccionismo: 15, 242

sadismo: 271
Salter, W. E. G.: 205
Sartre, J.-P.: 77, 178, 180, 253 *n*, 261, 271, 276, 278, 285, 286, 291, 293
satisfacer: 101, 224
Schafer, R.: 273, 286 *n*
Schelling, T. S.: 97 *n*, 162, 172
Schumpeter, J.: 18 *n*, 26, 29 *n*, 169, 172
Seducción: 140, 142
Sejersted, F.: 153, 159

selección: artificial, 57, 228; de grupo: 51, 240, 241; de parentesco: 51, 240, 241; natural: 12 *ss*, 30, 59, 226, 228, 240
Sen, A. K.: 42, 141
Simon, H.: 101, 103
sinceridad: 264, 290, 291
Stendhal: 70, 87, 214 *n*, 262, 264, 275, 277, 283
Stinchcombe A.: 54 *ss*, 86, 186 *n*
Strotz, R.: 66, 115, 123, 126, 152

Taylor, C.: 178, 179 *n*, 274, 293 *n*
Taylor: M.: 44 *n*, 113, 164 *n*, 209 *n*, 238, 243 *n*
tiempo: continuo vs. discreto: 119, 234; y funcionalismo: 63, 65; y amor: 277 *ss*; y preferencias: 115, 144; y racionalidad: 115, 244
Tisdell, C.: 105
tradición: 229 *ss*, 257
Trivers, R.: 23, 241
Valéry, P.: 87, 285, 291
voluntad; flaqueza de la: 79, 92 *n*, 118, 147, 188, 229, 288, 289
desear lo que no se puede desear: 251, 261
fuerza de: 73, 79, 174, 175

Watzlawick, P.: 251 *n*, 252 *n*, 266, 270
Weber, M.: 158, 197 *n*
Weiszäcker, C. von: 134 *ss*, 246 *n*
Williams, B. A. O.: 87, 94, 117 *n*, 251, 292
Williams, G. C.: 41 *n*, 62 *n*, 240 *n*
Wilson, E. O.: 45
Winter, S.: 18 *n*, 28 *n*, 103, 225
Yo: jerarquías de Yoes: 73, 185
Responsabilidad por el Yo: 178 *ss*, 275
Auto-engaño: 266, 285

ÍNDICE GENERAL

Prefacio y agradecimientos 7

I. La racionalidad perfecta: más allá del ascenso de la gradiente 11

1. Introducción 11

2. La máquina maximizadora local . . 16

3. La máquina maximizadora global . . 23

4. La conducta estratégica en los animales y el hombre 38

5. La explicación funcionalista en sociología 54

II. La racionalidad imperfecta: Ulises y las sirenas 66

1. Introducción 66

2. Hacia una definición 68

3. Pascal 84

4. Descartes 95

5. Preferencias incongruentes de tiempo 115

6. El cambio endógeno de preferencias . 132

7. El compromiso previo en la conducta animal 146

8. Abdicación del poder 150

9. Algunas conclusiones y nuevas preguntas 174

III. La racionalidad problemática: algunos problemas no resueltos en la teoría de la conducta racional 189

1. Introducción 189

2. Juegos sin soluciones 197

3. Preferencias lexicográficas . . 208

4. La probabilidad subjetiva . 214

5. Maximización, satisfacción y selección natural 223

6. La conducta tradicional y la conducta al azar 228

7. Explicación del altruismo 235

8. La inconstancia 244

9. La paradoja 248

10. ¿Y con eso qué? 254

IV. La irracionalidad: las contradicciones de la mente 260

1. Introducción 260
2. El odio 266
3. El amor 273
4. El autoengaño 285

Referencias 299

Índice analítico 319

Jon Elster

ULISES Y LAS SIRENAS

Situado en el cruce de disciplinas como la biología, la sociología, la literatura y las matemáticas, el autor hace una descripción y un análisis interpretativo de los comportamientos racional e irracional. Dejando de lado la analogía funcionalista, que pretendía explicar la conducta del hombre refiriéndola al sustrato biológico, la actitud racional —entendida como una capacidad para relacionarse con el futuro— se estabiliza para definir lo humano, no obstante los grados, imperfecciones y problemas que presenta en su ejecución concreta.

Hay una racionalidad imperfecta, ejemplificada con la fábula de Ulises y las Sirenas, que lleva al actor a atarse a un madero y denegar de sus potencias a causa de su débil voluntad. Contraria a la conducta racional, la irracional se apodera del hombre en el amor, el odio y el autoengaño. Actitudes como inconstancia y altruismo dan cuenta de otra modalidad del comportamiento, la racionalidad problemática, en la que tiene cabida ya una explicación por el azar.

Pascal, Descartes, John Donne; la filosofía, la poesía y las probabilidades sustentan este análisis abstracto y multifocal de las motivaciones internas que constituyen la esfera de la libertad. El punto de partida, las objeciones y las tesis propuestas tienen un soporte científico, un desarrollo riguroso y un máximo de orden y claridad en la exposición.